UNIVERSIDADE FEDERAL DO CEARÁ
CENTRO DE CIÊNCIAS
DEPARTAMENTO DE COMPUTAÇÃO
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA COMPUTAÇÃO

PAULO HENRIQUE MACÊDO DE ARAÚJO

THE GEODESIC CLASSIFICATION PROBLEM ON GRAPHS

FORTALEZA
2019

PAULO HENRIQUE MACÊDO DE ARAÚJO

THE GEODESIC CLASSIFICATION PROBLEM ON GRAPHS

Tese apresentada ao Programa de Pós-Graduação em Ciência da Computação do Centro de Ciências da Universidade Federal do Ceará, como requisito parcial à obtenção do título de doutor em Ciência da Computação. Área de Concentração: Algoritmos e Otimização.

Orientador: Prof. Dr. Manoel Bezerra Campêlo Neto

Coorientador: Prof. Dr. Ricardo Cordeiro Corrêa

FORTALEZA

2019

Dados Internacionais de Catalogação na Publicação
Universidade Federal do Ceará
Biblioteca Universitária
Gerada automaticamente pelo módulo Catalog, mediante os dados fornecidos pelo(a) autor(a)

A69t Araújo, Paulo Henrique Macêdo de.
The Geodesic Classification Problem on Graphs / Paulo Henrique Macêdo de Araújo. – 2019.
148 f. : il. color.

Tese (doutorado) – Universidade Federal do Ceará, Centro de Ciências, Programa de Pós-Graduação em Ciência da Computação , Fortaleza, 2019.
Orientação: Prof. Dr. Manoel Bezerra Campêlo Neto.
Coorientação: Prof. Dr. Ricardo Cordeiro Corrêa.

1. Classification. 2. Geodesic Convexity. 3. Polyhedral Combinatorics. I. Título.
CDD 005

PAULO HENRIQUE MACÊDO DE ARAÚJO

THE GEODESIC CLASSIFICATION PROBLEM ON GRAPHS

Tese apresentada ao Programa de Pós-Graduação em Ciência da Computação do Centro de Ciências da Universidade Federal do Ceará, como requisito parcial à obtenção do título de doutor em Ciência da Computação. Área de Concentração: Algoritmos e Otimização.

Aprovada em: 12 de Agosto de 2019

BANCA EXAMINADORA

Prof. Dr. Manoel Bezerra Campêlo Neto (Orientador)
Universidade Federal do Ceará (UFC)

Prof. Dr. Ricardo Cordeiro Corrêa (Coorientador)
Universidade Federal Rural do Rio de Janeiro (UFRRJ)

Prof. Dra. Martine Labbé
Université Libre de Bruxelles (ULB)

Prof. Dr. Javier Marenco
Universidade de Buenos Aires (UBA)

Prof. Dr. Rafael Castro de Andrade
Universidade Federal do Ceará (UFC)

I dedicate this work to my mother and father and to all those who contributed directly or indirectly with its realization.

ACKNOWLEDGEMENTS

I would like to thank my Ph.D. advisor Professor Dr. Manoel Campêlo for all his help and friendship. Without his excellent guidance, this thesis would not have been possible.

I would like to thank Professor Dra. Martine Labbé and my Ph.D. co-advisor Professor Dr. Ricardo Corrêa for all their very important contributions. Mainly, I thank Professor Dr. Ricardo Corrêa for guiding me since the initial steps of the academic area, and Professor Dra. Martine Labbé for receiving me in the Ph.D. internship.

Also, I would like to thank my committee members, Professor Dr. Rafael Castro de Andrade, Professor Dra. Yoshiko Wakabayashi and Professor Dr. Javier Marenco for all their contributions.

I would also like to thank my fiancée Raiza and all my friends from the ParGO research group.

This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Finance Code 001.

“Do not worry about the results of your actions, just pay attention to the action itself. The result will come from their own.”

(Eckhart Tolle)

RESUMO

Definimos e estudamos uma versão discreta do clássico problema de classificação no espaço Euclidiano. O problema em questão é definido em um grafo, onde os vértices não classificados precisam ser classificados levando em consideração a classificação dada para outros vértices. A partição de vértices em classes é baseada no conceito de convexidade geodésica em grafos, como uma substituta da convexidade Euclidiana no espaço multidimensional. Chamamo-lo de Problema de Classificação Geodésica - CG (Geodesic Classification Problem, em inglês) e consideramos duas variantes: duas classes, único grupo e duas classes, multigrupo. Propomos abordagens baseadas em programação inteira para cada versão considerada do problema CG, assim como um algoritmo de *branch-and-cut* para resolvê-las exatamente. Fizemos também um estudo dos poliedros associados, o que inclue a determinação de algumas famílias de desigualdades que definem facetas e algoritmos de separação. Condições para definição de facetas para a versão único grupo foram traduzidas para a versão multigrupo. Relacionamos nossos resultados com alguns já conhecidos na literatura para a classificação Euclidiana. Finalmente, realizamos experimentos computacionais para avaliar a eficiência computacional e a acurácia da classificação das abordagens propostas, comparando-as com alguns métodos de resolução clássicos para o problema de classificação com convexidade Euclidiana.

Palavras-chave: Classificação. Convexidade Geodésica. Combinatória Poliédrica.

ABSTRACT

We define and study a discrete version of the classical classification problem in the Euclidean space. The problem is defined on a graph, where the unclassified vertices have to be classified taking into account the given classification of other vertices. The vertex partition into classes is grounded on the concept of geodesic convexity on graphs, as a replacement for the Euclidean convexity in the multidimensional space. We name such a problem the Geodesic Classification (GC) problem and consider two variants: 2-class single-group and 2-class multi-group. We propose integer programming based approaches for each considered version of the GC problem along with branch-and-cut algorithms to solve them exactly. We also carry out a polyhedral study of the associated polyhedra, which includes some families of facet-defining inequalities and separation algorithms. Facetness conditions for the single-group case are carried over to the multi-group case. We relate our findings with results from the literature concerning Euclidean classification. Finally, we run computational experiments to evaluate the computational efficiency and the classification accuracy of the proposed approaches by comparing them with some classic solution methods for the Euclidean convexity classification problem.

Keywords: Classification. Geodesic Convexity. Polyhedral Combinatorics.

FIGURE LIST

Figure 1 – Example of the Euclidean classification problem in $\mathbb{R}^2$. Circles are points of class 1, while squares represent points of class 2. 31
Figure 2 – Example of Figure 1 in which some points are misclassified, becoming discrepant points. 32
Figure 3 – Equivalent conditions for the existence of linear separation. 33
Figure 4 – Example of a separating hyperplane and margins of the classes for the example of Figure 1. 34
Figure 5 – Examples of $\mathcal{N}$-set inequalities that define facets of P_E ((BLAUM *et al.*, 2019a)). 38
Figure 6 – An instance for the classification problem in which the sample set is linearly inseparable and poorly classified with a single separating hyperplane. 39
Figure 7 – Hyperplanes to separate each pair of groups of opposite classes for the example of Figure 6, with $L_1 = L_2 = 2$. 40
Figure 8 – An example of a classification graph. The filled vertices are those associated with the two classes. 43
Figure 9 – Vertex partition of a classification graph according to the geodesic convexity. 44
Figure 10 – An example of a classification graph where (V_B, V_R) is linearly inseparable and $(V_B \setminus \{v_4\}, V_R)$ is linearly separable. Filled circles are vertices of V_B, while squares are vertices of V_R. 45
Figure 11 – An example of solution with an outlier for the example of Figure 10. 45
Figure 12 – An example where $H[A_B] \cap H[A_R] \neq \emptyset$ but $H[A_B] \cap H[A_R] \cap V_N = \emptyset$ in the geodesic classification problem, for $A_B = V_B$ and $A_R = V_R \setminus \{v_3, v_4\}$. 47
Figure 13 – An example of a generalized 3-path constraint. 55
Figure 14 – An example of X-swing constraint. Vertex 5 belongs to V_N. 55
Figure 15 – C_4 inequality example. 57
Figure 16 – Example of a star tree inequality for ILP1. 58
Figure 17 – An example of alternating path inequality for ILP1. 61
Figure 18 – An example of star tree inequality for ILP2. 69
Figure 19 – An example of generalized convexity inequality for ILP2. 70
Figure 20 – Generalized walk inequality example. 74
Figure 21 – An example of instance for the geodesic classification problem. 76

Figure 22 – An example of solution with multi-group for the example of Figure 21. . . . 76
Figure 23 – Relations of the polyhedra. 113
Figure 24 – An illustration of corresponding solutions. 114
Figure 25 – Graph instance example. 119
Figure 26 – Separation algorithm for the example of Figure 25. Step 1. 119
Figure 27 – Separation algorithm for the example of Figure 25. Step 2. 119
Figure 28 – Separation algorithm for the example of Figure 25. Step 3. 119
Figure 29 – Running time versus density: Algorithm 3 and $br = 20$. 135
Figure 30 – Running time versus density: Algorithm 4 and $br = 20$. 136
Figure 31 – Running time versus density: Algorithm 3 and $br = 40$. 136
Figure 32 – Running time versus density: Algorithm 4 and $br = 40$. 136
Figure 33 – Running time versus density: Algorithm 3 and $br = 60$. 137
Figure 34 – Running time versus density: Algorithm 4 and $br = 60$. 137
Figure 35 – Running time versus density: Algorithm 3 and $br = 80$. 137
Figure 36 – Running time versus density: Algorithm 4 and $br = 80$. 138
Figure 37 – Synthetic instance example for $d = 2$. 139

TABLE LIST

Table 1 – Points in F_M. The first column indicates the group that the point belongs to. The remaining columns are entries of the point. E is the matrix of ones. '$\in F$' indicates that each row-vector is a point in F. 101
Table 2 – Pure integer linear model ILP2 and Algorithm 4 running times comparison for random instances with $br = 20$ (in seconds). 126
Table 3 – Pure integer linear model ILP2 and Algorithm 4 running times comparison for random instances with $br = 40$ (in seconds). 127
Table 4 – Pure integer linear model ILP2 and Algorithm 4 running times comparison for random instances with $br = 60$ (in seconds). 128
Table 5 – Pure integer linear model ILP2 and Algorithm 4 running times comparison for random instances with $br = 80$ (in seconds). 129
Table 6 – Algorithm 3 and Algorithm 4 running times comparison for random instances with $br = 20$ (in seconds). 130
Table 7 – Algorithm 3 and Algorithm 4 running times comparison for random instances with $br = 40$ (in seconds). 131
Table 8 – Algorithm 3 and Algorithm 4 running times comparison for random instances with $br = 60$ (in seconds). 132
Table 9 – Algorithm 3 and Algorithm 4 running times comparison for random instances with $br = 80$ (in seconds). 133
Table 10 – Effect of the valid inequalities for ILP1 for random instances. 133
Table 11 – Effect of the valid inequalities for ILP2 for random instances. 133
Table 12 – Algorithm 3, Algorithm 4, *SVM* and *MLP* comparison for Parkinson's instances with $br = 80\%$. 141
Table 13 – Algorithm 3, Algorithm 4, *SVM* and *MLP* comparison for Parkinson's instances with $br = 70\%$. 142
Table 14 – Algorithm 3, Algorithm 4, *SVM* and *MLP* comparison for *SPECT* instances with $br = 80\%$. 142
Table 15 – Algorithm 3, Algorithm 4, *SVM* and *MLP* comparison for *SPECT* instances with $br = 70\%$. 142
Table 16 – Algorithm 3, Algorithm 4, *SVM* and *MLP* comparison for synthetic instances with dimension 2 and $br = 80\%$. 143

Table 17 – Algorithm 3, Algorithm 4, *SVM* and *MLP* comparison for synthetic instances with dimension 2 and $br = 70\%$. 143
Table 18 – Algorithm 3, Algorithm 4, *SVM* and *MLP* comparison for synthetic instances with dimension 3 and $br = 80\%$. 144
Table 19 – Algorithm 3, Algorithm 4, *SVM* and *MLP* comparison for synthetic instances with dimension 3 and $br = 70\%$. 144

SYMBOL LIST

V_B	Set of initially classified blue vertices
V_R	Set of initially classified red vertices
V_{BR}	Set of all initially classified vertices
C_B	Set of indices of the blue groups
C_R	Set of indices of the red groups
C_{BR}	Set of indices of all groups
$K(i)$	Class of the initially classified vertex i
$\bar{K}(i)$	Opposite class of the initially classified vertex i
ILP1	Set covering formulation
ILP2	A more compact formulation
ILP1^M	Set covering formulation (multi-group)
ILP2^M	A more compact formulation (multi-group)
ILP3^M	A even more compact formulation (multi-group)

SUMMARY

1 INTRODUCTION . . . 17

1.1 Applications . . . 19

1.2 Results and contributions . . . 19

1.3 Text structure . . . 20

2 PRELIMINARIES . . . 21

2.1 Graph theory . . . 21

2.2 Basic concepts of polyhedra and linear programming . . . 23

2.2.1 Affine transformations . . . 25

2.2.2 Projection . . . 27

2.3 The set covering problem . . . 29

3 EUCLIDEAN CLASSIFICATION PROBLEM . . . 31

3.1 Linear separation . . . 32

3.2 Outlier points . . . 35

3.2.1 $\mathscr{N}$-Set inequalities . . . 37

3.3 Piecewise linear separation . . . 39

4 GEODESIC CLASSIFICATION PROBLEM . . . 43

4.1 Integer formulations for the 2-SGC problem . . . 47

4.1.1 A set covering formulation for the 2-SGC problem . . . 47

4.1.2 ILP1's polyhedron . . . 49

4.2 $\mathscr{N}$-set inequalities . . . 50

4.2.1 ILP1's constraints . . . 53

4.2.2 Valid and facet-defining $\mathscr{N}$-set non-elementary inequalities for P_1 . . . 55

4.2.3 A compact formulation for the 2-SGC problem . . . 61

4.2.4 ILP2's polyhedron . . . 62

4.2.5 ILP2's constraints . . . 64

4.2.6 Valid and facet-defining inequalities for P_2 . . . 67

5 PIECEWISE LINEAR SEPARATION FOR THE GEODESIC CLASSIFICATION PROBLEM . . . 76

5.1 Computational complexity of the 2-MGC problem . . . 78

5.2 Integer formulations for the 2-MGC problem . . . 79

5.2.1 A set covering formulation for the 2-MGC problem . . . 80

5.2.2	*The associated polytope - P_1^M*	81
5.2.3	*Relations with P_1*	82
5.2.4	*Valid inequalities and facets*	86
5.2.5	*A compact formulation for the 2-MGC problem*	92
5.2.6	*The associated polytope - P_2^M*	95
5.2.7	*Relations with P_1^M*	95
5.2.8	*Relations with P_2*	96
5.2.9	*Valid inequalities and facets*	99
5.3	**A more compact formulation for the 2-MGC problem**	103
5.3.1	*The associated polytope - P_3^M*	104
5.3.2	*Relations with P_2^M*	105
5.3.3	*Valid inequalities and facets*	107
5.3.4	*Summary of facetness results*	110
5.3.5	*Comparison of the formulations*	113
6	**ALGORITHMS AND COMPUTATIONAL EXPERIMENTS**	116
6.1	**Calculation of D_{hj}**	116
6.2	**Convex hull calculation**	116
6.3	**Pre-processing**	117
6.4	**Separation algorithms**	118
6.4.1	*C_4 inequalities separator*	118
6.4.2	*Generalized convexity inequalities separator*	118
6.4.3	*Generalized alternating path inequalities separator*	122
6.4.4	*Lazy constraints separator for elementary $\mathcal{N}$-set inequalities*	122
6.4.5	*Lazy constraints separator for convexity constraints*	123
6.5	**Geodesic classification algorithms**	123
6.6	**Results and analysis**	125
6.6.1	*Random instances*	125
6.6.2	*Realistic and synthetic instances*	137
6.6.3	**SVM** *algorithm*	139
6.6.4	**Neural networks** *algorithm MLP*	140
6.6.5	*Parkinson's disease instances*	140
6.6.6	*SPECT heart data instances*	140

6.6.7 ***Analysis of the realistic instances experiments*** 141
6.6.8 ***Analysis of the synthetic instances experiments*** 143
7 **CONCLUDING REMARKS** 145
REFERENCES 147

1 INTRODUCTION

Supervised learning denotes the automatic prediction of the behavior of unknown data based on a set of samples. It is a tool widely used in many everyday situations of the nowadays information society. In general terms, it can be described by the following two-phase procedure: in the initial phase, or *training phase*, the sample set is analyzed. Each sample consists of an array of encoded attributes that characterize an object of a certain type together with a label that associates a class to the corresponding object. Most commonly, only two classes are considered. A tacit assumption made at this phase is that there is an underlying pattern associated with the samples of each class that sets them apart from the samples of the other classes. Thus, the purpose of the training phase is to determine a mapping from all possible objects into the set of possible classes as an extension of an underlying pattern of the samples. Then, in the second phase, the mapping determined in the training phase is used to respond to queries for the class of objects that do not belong to the sample set.

An optimization problem is usually associated with the training phase. Referred to as *classification problem*, it consists in grouping similar samples to get clusters as internally homogeneous as possible. A wide range of solution methods is available, each depending on the coding of the samples and the criterion adopted to express homogeneity. A prevalent approach is to encode the samples as vectors in the Euclidean space and to assume that the class patterns can be appropriately characterized by convex sets. In this vein, continuous optimization methods, including linear and quadratic programming, have been developed in the last 40 years. See e.g. (ARTHANARI; DODGE, 1993; CARRIZOSA; MORALES, 2013; CORTES; VAPNIK, 1995; FREED; GLOVER, 2007; PARDALOS; HANSEN, 2008). More recently, integer linear programming tools started to be used in conjunction with continuous methods, as we can see in (BERTSIMAS; SHIODA, 2007; MASKOOKI, 2013; SUN, 2011; UNEY; TURKAY, 2006; XU; PAPAGEORGIOU, 2009).

Inspired by the version of the classification problem based on Euclidean convexity discussed in (CORRÊA *et al.*, 2019), we define a new variant of the classification problem that is stated in terms of notions of convexity in graphs. For this new problem, we develop some integer linear programming formulations. The main purpose of this thesis is the structural study of the polyhedra associated with these formulations, which can help make the Euclidean solution methods more robust. The statement of this classification problem assumes the following hypotheses:

1. The objects are not encoded numerically. Instead, each object is characterized by its similarities with other objects. The configuration of the objects is thus represented by a similarity graph $G = (V,E)$, connected, where V is the set of all objects, and E gives the pairs of similar objects. The objects associated with the sample set constitute a proper subset of V. In addition, it is assumed the existence of an underlying sample patterns that can be expressed, or at least approximated, by the notion of *geodesic convexity* in graphs ((PELAYO, 2013)). Such a convexity is defined with respect to the shortest paths in G (analogously to the definition of Euclidean convexity with respect to the Euclidean distances between points in $\mathbb{R}^n$).
2. The sample set may contain an arbitrary number of misclassified objects, called *outliers*, which result from possible sampling errors or due to inherent characteristics of the phenomenon being modeled. From the mathematical point of view, an outlier is that classified object that leads the underlying pattern of the samples in its class to deviate from the convexity definition. The possible occurrence of outliers poses an additional challenge to any method used to solve the classification problem since they have to be detected and disregarded so that accurate solution may be found. The goal is to divide the vertex set into 2 classes, based on the classification of the samples. Each class may comprise a single group of vertices, in the basic version of the problem, or multi-group of vertices, in the generalized version.

Considering the hypotheses above, the 2-*class Single-group Geodesic Classification* (2-SGC) problem and the 2-*class Multi-group Geodesic Classification* (2-MGC) problem tackled in this thesis become purely combinatorial. In integer linear programming formulations, those combinatorial aspects are expressed by binary variables used for two purposes:

1. Identification, and possibly counting, of points considered to be discrepant. Once identified, the outliers can be disregarded in the obtaining of the patterns that yield a solution to the classification problem. In several formulations proposed in the literature, the optimization criterion is the minimization of the number of outlier points, which implies the necessity of counting such points.
2. Division of the object sets (vertices) into subsets and their associations to classes.

From the practical point of view, these problems allow encoding object similarities through some binary relation (instead of Euclidean distances), a fact that benefits many practical applications in big data. As detailed later, a solution is neither a covering nor a partitioning of G

in convex sets in the sense studied in (ARTIGAS *et al.*, 2011; BUZATU; CATARANCIUC, 2015). Thus, besides the applications, the study of the geodesic classification problems is motivated by theoretical interest since it brings with it the possibility of establishing new interesting problems on graphs. Moreover, these problems can be seen as the combination of a graph convexity problem and the well known set covering problem ((KARP, 1972)), as shown by the mathematical models proposed in Sections 4.1 and 5.2. Besides, the study of their combinatorial structure may be useful to design solution methods for other versions of the classification problem, including those based on Euclidean convexity.

To the best of our knowledge, these geodesic classification problems have not been mentioned in the literature yet. We state three integer programming formulations and derive some families of facet-defining inequalities (a part of this work can be seen in (ARAÚJO *et al.*, 2019) where another integer linear formulation is used). In addition, we present a branch-and-cut algorithm for each formulation and run some computational experiments. The accuracy of the geodesic convexity approach is validated by comparing the prediction provided by the proposed algorithm with the one obtained, for similar instances, by *SVM* and *MLP* (these are two of the most used approaches for the Euclidean convexity classification problem).

1.1 Applications

Applications for the geodesic classification problem are easily found in the fields of data mining and classical statistics. As examples, we have text classification and communities detection in social networks, historic files similarity prediction, content recommendation in Netflix, and spam filtering for e-mails. In the text classification on Twitter, for instance, we want to find text mining tools that help us to understand messages on Twitter, as for sentimental analysis, like recommendation, friend recommendation and others ((HONG; DAVISON, 2010)). In these contexts, combinatorial aspects emerge from the relations between samples. The same can be applied to many other social network applications.

1.2 Results and contributions

We introduce two versions of a classification pattern defined on graphs using the notion of geodesic convexity. Such a notion is formalized in the definition of two versions of the problem: 2-*class Single-group Geodesic Classification* (2-SGC) problem and 2-*class Multi-*

group Geodesic Classification (2-MGC) problem, which have not appeared in the literature. The division of the problem into these two versions allows us to study the inherent properties of the different aspects of the classification approach. Inspired by the work presented by (CORRÊA *et al.*, 2019) on Euclidean classification, we present three integer linear programming formulations to solve such problems. The first one is a set covering formulation with an exponential number of constraints, while the second one is a compact formulation with a polynomial number of constraints. We carry out some theoretic comparison between all formulations and study the associated polyhedra, giving some valid inequalities, families of facet-defining inequalities and separation algorithms. Results for the 2-SGC problem are carried over to the 2-MGC problem. Finally, we show a branch-and-cut algorithm to solve each integer formulation and run computational experiments to compare the proposed approaches. We analyze two aspects: the computational performance of the solution methods and the accuracy of the generated solutions.

1.3 Text structure

This text is organized as follows. Chapter 2 presents some basic concepts notation and results in graph theory, linear algebra, polyhedral combinatorics and linear programming. In Chapter 3, we formalize the Euclidean convexity classification problem and show some solution methods found in the literature. The definition of the 2-class single-group geodesic classification problem is introduced in Chapter 4 along with integer linear formulations and a study of the associated polyhedra. Analogously, the definition of the 2-class multi-group geodesic classification problem and integer linear formulations are presented in Chapter 5. In Chapter 6, we present the branch-and-cut algorithms and show computational experiments results to evaluate the performance of each approach. Finally, we present some concluding remarks and directions for future works in Chapter 7.

2 PRELIMINARIES

In this chapter, we show some basic concepts and properties from graph theory, linear algebra, polyhedral combinatorics and linear programming. We also establish the notation. Most of the statements are well known so that proofs are usually omitted. However, in subsections 2.2.1 and 2.2.2, we present some results that will be frequently used in Chapters 4 and 5 and that are not easily found in a text book.

2.1 Graph theory

Basic concepts in graph theory are easily found in any introductory graph book. In particular, we follow the terminology used in (BONDY; MURTY, 2008).

A *simple graph* G is an ordered pair (V,E), where V is a finite set of elements called *vertices* and E is a set of elements called *edges*, with each edge being a non-ordered pair of distinct vertices (when they are ordered pair of vertices, we call them *arcs* and the graph is called a *directed graph*. It is called *dag* or *direct acyclic graph* if there is no cycle). We denote by $V(G)$ the set of vertices of G and $E(G)$ (resp., $A(G)$, in case of arcs) the set of edges (resp., arcs) of G.

If $e = \{u,v\}$ is an edge, then we say that *e affects u* and *v*, or *u* and *v* are its *extreme points* or that *u* and *v* are *adjacent*. We also denote $\{u,v\}$ by uv.

A *walk* in a graph G is a sequence of vertices $W = (v_1,v_2,\dots,v_l)$, $l \geq 1$, such that $\{v_i,v_{i+1}\} \in E(G)$ for $i = 1,\dots,l-1$. In this case, we call $V(W) = \{v_1,\dots,v_l\}$ and $E(W) = \{\{v_i,v_{i+1}\} : i = 1,\dots,l-1\}$. We say that v_1 is the *start* or *source*, and v_l is the *end* of W, and that W is a walk from v_1 to v_l. The *length* of W is $l-1$. If W has no repeated vertices, then we say that W is a *path* (between v_1 and v_l). In this case, its length, $l-1$, is equal to the number of edges or, equivalently, to the number of vertices minus 1. When this length is minimum among all paths between v_1 and v_l in G, we say that it is a shortest path from v_1 to v_l. If $P = (v_1,v_2,\dots,v_l)$ is a path with $l \geq 3$ and $\{v_l,v_1\} \in E(G)$, then $C = (v_1,v_2,\dots,v_l,v_1)$ is a *cycle*.

An *incomplete path* (resp., *incomplete walk*) P_I from v_1 to v_l in a graph G is a subsequence, not necessarily continuous, of a path (resp., walk) between v_1 and v_l in G that contains v_1 and v_l. If a corresponding path of P_I is a shortest path in G, then we call P_I an *incomplete shortest path*. Sometimes, to distinguish from an incomplete path (resp., incomplete walk), we denote path (resp., walk) and shortest path by *complete path* (resp., complete walk) and *complete shortest path*, respectively.

A graph G is *connected* if, for any pair of distinct vertices u and v in G, there is at least one path between u and v. A *tree* is a connected graph with no cycle.

If two graphs $G = (V, E)$ and $H = (W, F)$ are such that $V \subseteq W$ and $E \subseteq F$, then G is called a *subgraph* of H. If $X \subseteq V(G)$, then the subgraph of G induced by X, denoted by $G[X]$, is the graph whose set of vertices is X and where edges are those of G with both extreme points in X. If G is a path, then a subgraph of G that is a path is called *subpath*.

A set of vertices S is an *independent set* if there is no edge between any pair of vertices of S, and it is a *clique* if there is an edge between any pair of vertices of S.

A *geodesic* between two vertices h and j in G is a shortest path between h and j in the graph and its length is denoted by $\delta(h, j)$. The closed interval $D[h, j]$ is the set of all vertices lying on a geodesic between h and j. We also denote $D_{hj} = D[h, j] \setminus \{h, j\} = \{i \in V \setminus \{h, j\} \mid$ i belongs to a shortest path between h and j$\}$. Given a set $S \subseteq V(G)$, $D[S] = \bigcup_{u,v \in S} D[u, v]$. If $D[S] = S$, then S is a *convex set*. For $k \geq 1$, let $D^k[S]$ be the result of the iterative application of operator D from S for k iterations, i.e. $D^1[S] = D[S]$ and $D^{k+1}[S] = D[D^k[S]]$. Note that $D^{k+1}[S] = D^k[S]$ if and only if $D^k[S]$ is convex. The *convex hull* of S, denoted by $H[S]$, is the smallest convex set containing S. This minimum set is unique. If $H[S] = V$, then S is a *hull set*. Observe that $H[S] = D^k[S]$ for some $k \geq 1$. In other terms, $H[S]$ can be obtained by the iterative application of D.

From (ARTIGAS *et al.*, 2011), the analogy between the concept of a convex set in continuous and discrete mathematics can be made by considering the vertex set of a connected graph and the distance between two vertices (number of edges in a shortest path between them) as a metric space. Thus, a vertex subset S of $V(G)$ is said to be convex if it contains the vertices of all shortest paths connecting any pair of vertices in S. This concept of convexity is called *geodesic convexity*. Other definitions of convexity have been studied by considering different path types such as chordless paths ((FARBER; JAMISON, 1986)) and triangle paths ((CHANGAT; MATHEW, 1999)).

Some of the early papers that generalize the Euclidean concepts of convex sets to graphs date to the eighties: (HARARY; NIEMINEN, 1981), (EDELMAN; JAMISON, 1985), (FARBER; JAMISON, 1986). More recently, convexity on graphs have been studied under several aspects, like geodesic sets, hull and convexity numbers ((CáCERES *et al.*, 2006), (DOURADO *et al.*, 2009a), (DOURADO *et al.*, 2010)).

2.2 Basic concepts of polyhedra and linear programming

A *linear programming problem* (LP) is a problem of minimizing (or maximizing) a linear function subject to a set of linear constraints. These constraints can be expressed as equalities and inequalities.

A *polyhedron* P is a set of the form $P = \{x \in \mathbb{R}^n : Ax \leq b\}$, where (A, b) is a $(m \times (n+1))$ matrix. In this case, we also denote P by $P(A, b)$. Thus, LP can be seen as a problem of minimizing a linear function over a polyhedron.

An *integer linear programming problem* (ILP) is an LP where all variables can only receive integer values. When those integer values are restricted to 0 or 1, we call an ILP/01. While LP's can be solved in polynomial time, ILP's are NP-hard in general.

Formally, an integer linear programming is an LP in the form

$$\begin{aligned} \min \quad & c^T x \\ \text{s.t.} \quad & Ax \leq b \\ & x \in \mathbb{Z}^n_+. \end{aligned}$$

Every feasible solution of an ILP with a minimization objective function gives an *upper bound* for the optimal value, where $f(x) = c^T x$ is called the objective function and every point $x \in \mathbb{Z}^n_+$ such that $Ax \leq b$ is called a feasible solution. However, there are problems in which finding good feasible solutions is as hard as solving the ILP itself. On the other hand, there are polynomial methods to find lower bounds for an ILP. One of these methods consists in solving the problem obtained by removing the integrality constraints, which is called the *linear relaxation* of the ILP.

In general, a problem $z^R = min\{f(x) : x \in T \subseteq \mathbb{R}^n\}$ is a *relaxation* of $z = min\{f(x) : x \in X \subseteq \mathbb{Z}^n\}$ if $X \subseteq T$ and $f(x) \leq c(x)$ for each $x \in X$. Observe that $z \geq z^R$.

A point $x \in \mathbb{R}^n$ is a *linear combination* of points $x_1, \ldots, x_l \in \mathbb{R}^n$ if, for some $\alpha = (\alpha_1, \ldots, \alpha_l) \in \mathbb{R}^l$, $x = \sum_{i=1}^{l} \alpha_i x_i$. Such a linear combination is called

1. *affine combination* if $\alpha_1 + \alpha_2 + \ldots + \alpha_l = 1$;
2. *conic combination* if $\alpha_1, \alpha_2, \ldots, \alpha_l \geq 0$;
3. *convex combination* if it is affine and conic.

For a non-empty set $S \subseteq \mathbb{R}^n$, we define the *affine hull* of the elements of S, denoted by affine(S), as the set of all points that are an affine combination of a finite number of elements of S. Analogously, we define the *convex hull* of S, denoted by *conv*(S), as the set of all points that are a convex combination of a finite number of points of S.

A set $S \subseteq \mathbb{R}^n$ is *linearly independent* if, for any finite subset of points $\{x_1, \ldots, x_l\}$ of S and $\alpha \in \mathbb{R}^l$ such that $\sum_{i=1}^{l} \alpha_i x_i = 0$, we have $\alpha_1 = \alpha_2 = \ldots = \alpha_l = 0$. Similarly, $S \subseteq \mathbb{R}^n$ is *affinely independent* if, for any finite subset $\{x_1, \ldots, x_l\}$ of S and $\alpha \in \mathbb{R}^l$ such that $\sum_{i=1}^{l} \alpha_i x_i = 0$ and $\sum_{i=1}^{l} \alpha_i = 0$, it hold that $\alpha_1 = \alpha_2 = \ldots = \alpha_l = 0$. For some proofs of affinely independence along this text, we use the notation $e^i \in \{0,1\}^n$ to mean the binary vector with value 1 only in the entry indexed by i. Besides, we denote $e = \sum_i e^i$.

For $S \subseteq \mathbb{R}^n$, the *rank* of S, denoted by *rank*(S), is the cardinality of a largest subset of S that is linearly independent. Similarly, *affine-rank* of S, denoted by affine-rank(S), is the cardinality of a largest affinely independent set contained in S.

The *dimension* of a polyhedron P, denoted by $dim(P)$, is the maximum number of affinely independent points in P minus one, i.e., $dim(P) =$*affine-rank*$(P) - 1$. A polyhedron $P \subseteq \mathbb{R}^n$ has full dimension if $dim(P) = n$. It can be shown that, if affine$(P) = \{x \in \mathbb{R}^n : Ax = b\}$, then $dim(P) = n - rank(A)$.

We say that an inequality $a^T x \leq a_0$ is a *valid inequality* for a set S if $a^T w \leq a_0$ for all $w \in S$.

We call F a face of polyhedron P if $F = \{x \in P : a^T x = a_0\}$ for a valid inequality $a^T x \leq a_0$ for P. A non-empty face F of P is a *facet* if $dim(F) = dim(P) - 1$. If F is a facet of P and $F = \{x \in P : \gamma^T x = \gamma_0\}$, where $\gamma^T x \leq \gamma_0$ is valid for P, we say that $\gamma^T x \leq \gamma_0$ *defines* (or *induces*) the facet F.

If P has full dimension, then P has a unique minimal description, given by the inequalities that define the facets of P. By minimal we mean that removing any of these inequalities yields a different polyhedron. The uniqueness is implied by the fact that every facet-defining inequality of a full -dimensional polyhedron has a unique expression, except for scalar multipliers. When the polyhedron does not have full dimension, the facet-defining inequalities are not expressed in a unique form. Thus, showing that an inequality defines a facet usually demands a more laborious proof in the non-full dimensional case. Fortunately, all polyhedra studied in this work are full-dimensional.

2.2.1 Affine transformations

An affine transformation is a mapping $\mathscr{Q} : \mathbb{R}^n \to \mathbb{R}^m$ such that $\mathscr{Q}(x) = Qx + q$ for some matrix $Q \in \mathbb{R}^{m \times n}$ and vector $q \in \mathbb{R}^m$. Injectiveness and surjectiveness of $\mathscr{Q}(x) = Qx + q$ are given by the rank of Q (ROCKAFELLAR, 1997). Given a set $P \subseteq \mathbb{R}^n$, let $\mathscr{Q}(P) = \{\mathscr{Q}(x) : x \in P\}$. Valid inequalities for P and $\mathscr{Q}(P)$ can be related as follows.

Proposition 2.2.1 *Let $Q \in \mathbb{R}^{m \times n}$. If $\operatorname{rank}(Q) = n$, then $LQ = I$ for $L = (Q^T Q)^{-1} Q^T$. If $\operatorname{rank}(Q) = m$, then $QR = I$ for $R = Q^T (QQ^T)^{-1}$.*

Proof Suppose that $\operatorname{rank}(Q) = n$. To prove that $Q^T Q$ is invertible, we consider $x \in \mathbb{R}^n$ satisfying $Q^T Qx = 0$ and show that $x = 0$. Indeed, we have that $x^T Q^T Qx = 0$, that is, $(Qx)^T (Qx) = 0$, which leads to $Qx = 0$. Since $\operatorname{rank}(Q) = n$, we must have $x = 0$. Thus, $L = (Q^T Q)^{-1} Q^T$ is well-defined and clearly $LQ = (Q^T Q)^{-1} Q^T Q = I$. The second part is the first one for Q^T. ■

Proposition 2.2.2 *Let $\mathscr{Q} : \mathbb{R}^n \to \mathbb{R}^m$ be an affine transformation such that $\mathscr{Q}(x) = Qx + q$. If $\operatorname{rank}(Q) = n$, then $\mathscr{Q}$ is injective. If $\operatorname{rank}(Q) = m$, then $\mathscr{Q}$ is surjective.*

Proof Suppose that $\operatorname{rank}(Q) = n$. Let $x, x' \in \mathbb{R}^n$ such that $\mathscr{Q}(x) = \mathscr{Q}(x')$, i.e. $Qx = Qx'$. Using matrix L given by Proposition 2.2.1, we get $x = x'$. Then $\mathscr{Q}$ is injective. Now, suppose that $\operatorname{rank}(Q) = m$ and let matrix R be given by Proposition 2.2.1. Let $y \in \mathbb{R}^m$ and $x = R(y - q)$. Then, $\mathscr{Q}(x) = QR(y - q) + q = y$. Therefore, $\mathscr{Q}$ is surjective. ■

Proposition 2.2.3 *Let $P \subseteq \mathbb{R}^n$, and $\mathscr{Q} : \mathbb{R}^n \to \mathbb{R}^m$ and $\mathscr{R} : \mathbb{R}^m \to \mathbb{R}^n$ be two mappings. If $\mathscr{R}$ is affine, $\mathscr{R}(\mathscr{Q}(x)) = x$ for all $x \in P$ and $\pi^T x \leq \pi_0$ is valid for P, then $\pi^T \mathscr{R}(y) \leq \pi_0$ is a valid (linear) inequality for $\mathscr{Q}(P)$. Conversely, if $\mathscr{Q}$ is affine and $\mu^T y \leq \mu_0$ is valid for $\mathscr{Q}(P)$, then $\mu^T \mathscr{Q}(x) \leq \mu_0$ is a valid (linear) inequality for P.*

Proof Let $\bar{y} \in \mathscr{Q}(P)$. Then, there is $\bar{x} \in P$ such that $\bar{y} = \mathscr{Q}(\bar{x})$, and so $\mathscr{R}(\bar{y}) = \bar{x}$. Since $\pi^T \bar{x} \leq \pi_0$, we get $\pi^T \mathscr{R}(\bar{y}) \leq \pi_0$. Now, let $\bar{x} \in P$ and $\bar{y} = \mathscr{Q}(\bar{x}) \in \mathscr{Q}(P)$. Since $\mu^T \bar{y} \leq \mu_0$, we have that $\mu^T \mathscr{Q}(\bar{x}) \leq \mu_0$. Observe that $\pi^T \mathscr{R}(\bar{y}) \leq \pi_0$ and $\mu^T \mathscr{Q}(\bar{x}) \leq \mu_0$ are linear inequalities since $\mathscr{R}$ and $\mathscr{Q}$ are affine. ■

The statement of Proposition 2.2.3 can be rephrased in terms of the matrices defining $\mathscr{R}$ and $\mathscr{Q}$.

Proposition 2.2.4 *Let $P \subseteq \mathbb{R}^n$ and $\mathscr{Q} : \mathbb{R}^n \to \mathbb{R}^m$ be an affine transformation such that $\mathscr{Q}(x) = Qx + q$. If $\operatorname{rank}(Q) = n$ and $\pi^T x \leq \pi_0$ is valid for P, then $\pi^T Ly \leq \pi_0 + \pi^T Lq$ is valid for $\mathscr{Q}(P)$, where $L = (Q^T Q)^{-1} Q^T$. If $\mu^T y \leq \mu_0$ is valid for $\mathscr{Q}(P)$, then $\mu^T Qx \leq \mu_0 - \mu^T q$.*

Proof Assume that $\operatorname{rank}(Q) = n$ so that L is well-defined. Let $\mathscr{R} : \mathbb{R}^m \to \mathbb{R}^n$ be the affine transformation $\mathscr{R}(y) = Ly - Lq$. Then, $\mathscr{R}(\mathscr{Q}(x)) = L(Qx + q) - Lq = LQx = x$. By Proposition 2.2.3, we deduce that $\pi^T Ly \leq \pi_0 + \pi^T Lq$ is valid for $\mathscr{Q}(P)$. The second part is a direct consequence of Proposition 2.2.3. ■

We can also relate affinely independent points in P and $\mathscr{Q}(P)$.

Proposition 2.2.5 *Let $P \subseteq \mathbb{R}^n$ be a finite set, and $\mathscr{Q} : \mathbb{R}^n \to \mathbb{R}^m$ and $\mathscr{R} : \mathbb{R}^m \to \mathbb{R}^n$ be two mappings. If $\mathscr{R}$ is affine, $\mathscr{R}(\mathscr{Q}(x)) = x$ for all $x \in P$ and P is an affinely independent set, then $\mathscr{Q}(P)$ is affinely independent. Conversely, if $\mathscr{Q}$ is affine and $\mathscr{Q}(P)$ is affinely independent, then P is affinely independent.*

Proof Let $P = \{x^1, \ldots, x^p\}$ and $y^i = \mathscr{Q}(x^i)$ for $i = 1, ..., p$. Consider $\alpha_i \in \mathbb{R}$, $i = 1, ..., p$, such that $\sum_{i=1}^p \alpha_i = 0$. First, suppose that $\mathscr{R}$ is affine, $\mathscr{R}(y^i) = x^i$ and P is affinely independent. Then, $\sum_{i=1}^p \alpha_i y^i = 0$ implies $0 = \mathscr{R}\left(\sum_{i=1}^p \alpha_i y^i\right) - \mathscr{R}(0) = \sum_{i=1}^p \alpha_i \mathscr{R}(y^i) = \sum_{i=1}^p \alpha_i x^i$, where the second equality is implied by $\sum_{i=1}^p \alpha_i = 0$. As P is affinely independent, it must be $\alpha_i = 0$ for all $i = 1, ..., p$. Therefore, $\mathscr{Q}(P)$ is affinely independent. Now, suppose that $\mathscr{Q}$ is affine and $\mathscr{Q}(P)$ is affinely independent. Then, $\sum_{i=1}^p \alpha_i x^i = 0$ implies $0 = \mathscr{Q}\left(\sum_{i=1}^p \alpha_i x^i\right) - \mathscr{Q}(0) = \sum_{i=1}^p \alpha_i \mathscr{Q}(x^i)$. It follows that $\alpha_i = 0$ for all $i = 1, ..., p$, and so P is affinely independent. ■

Particularly, an affine transformation applied to a polyhedron results in a polyhedron.

Theorem 2.2.6 *If $P \subseteq \mathbb{R}^n$ is a polyhedron and $\mathscr{Q} : \mathbb{R}^n \to \mathbb{R}^m$ is an affine transformation, then $\mathscr{Q}(P) \subseteq \mathbb{R}^m$ is a polyhedron.*

In our context, however, we will be more interested in the application of an affine transformation to a subset of a polyhedron.

Proposition 2.2.7 *Let $P' = \{x \in \mathbb{R}^n : Ax \leq b\}$ and $\mathscr{Q} : \mathbb{R}^n \to \mathbb{R}^m$ be an affine transformation such that $\mathscr{Q}(x) = Qx + q$. If $\operatorname{rank}(Q) = n$ and $P \subseteq P'$, then $\mathscr{Q}(P) \subseteq \{y \in \mathbb{R}^m : ALy \leq b + ALq\}$, where $L = (Q^TQ)^{-1}Q^T$. If $\operatorname{rank}(Q) = m$ and $P \supseteq P'$ then $\mathscr{Q}(P) \supseteq \{y \in \mathbb{R}^m : ARy \leq b + ARq\}$, where $R = Q^T(QQ^T)^{-1}$.*

Proof If $\operatorname{rank}(Q) = n$ and $P' \subseteq P$, the result directly follows from Proposition 2.2.4. Now, suppose that $\operatorname{rank}(Q) = m$ and $P \supseteq P'$. By Proposition 2.2.1, $QR = I$ for $R = Q^T(Q^TQ)^{-1}$. Let $y \in \mathbb{R}^m$ such that $(AR)y \leq b + (AR)q$, that is, $AR(y - q) \leq b$. Then $x = R(y - q) \in P' \subseteq P$, and $Qx = y - q$. Therefore, $y \in \mathscr{Q}(P)$. ■

The composition of an affine mapping and the convexification operator is interchangeable.

Proposition 2.2.8 *Let $P \subseteq \mathbb{R}^n$ and $\mathscr{Q} : \mathbb{R}^n \to \mathbb{R}^m$ be an affine transformation. Then, $\mathscr{Q}(\operatorname{conv}(P))$ $= \operatorname{conv}(\mathscr{Q}(P))$.*

Proof Let I be a finite set and, for all $i \in I$, $x^i \in P$ and $\alpha_i \geq 0$ with $\sum_{i \in I} \alpha_i = 1$. Since $\sum_{i \in I} \alpha_i = 1$, we get that $\mathscr{Q}(\sum_{i \in I} \alpha_i x^i) = \sum_{i \in I} \alpha_i \mathscr{Q}(x^i)$. This equality leads to the desired result. ■

2.2.2 Projection

Let $Z \subseteq \mathbb{R}^p \times \mathbb{R}^q$ be a set where each point is given by a pair $(x, y) \in \mathbb{R}^p \times \mathbb{R}^q$. The (orthogonal) projection of Z onto the x-space is the set

$$\operatorname{proj}_x(Z) = \{x \in \mathbb{R}^p : (x, y) \in Z \text{ for some } y \in \mathbb{R}^q\}.$$

The projection operator is a special case of an affine (actually linear) transformation. Indeed, $\operatorname{proj}_x(Z) = \mathscr{Q}(Z)$, where $\mathscr{Q}(x, y) = \begin{bmatrix} I & 0 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix}$. Therefore, if Z is a polyhedron, then so is $\operatorname{proj}_x(Z)$. Moreover, the projected polyhedron can be described as follows.

Theorem 2.2.9 *Let $Z = \{(x,y) \in \mathbb{R}^p \times \mathbb{R}^q : Ax + By \leq b\}$ be a polyhedron. Then* $\text{proj}_x(Z) = \{x \in \mathbb{R}^p : (u^T A)x \leq u^T b, \forall u$ *extreme direction of* $U\}$*, where $U = \{u \geq 0 : u^T B = 0\}$ is called projection cone.*

Proof By definition, $x \in \text{proj}_x(Z)$ if, and only if, the set $\{y \in \mathbb{R}^q : By \leq b - Ax\}$ is nonempty. By Farkas's Lemma, it is equivalent to ask that $u^T(b - Ax) \geq 0$, for all $u \in U$. As U is a polyhedral cone, it suffices to consider the subset of these inequalities related to the extreme directions of U. ■

It is worth to mention that not every extreme direction of the projection cone induces a facet of $\text{proj}_x(Z)$. (BALAS, 1998) introduced a transformation that produces another cone whose extreme directions induce facets of the projection. Regardless, it is not an easy task to relate facets of the original and projected polyhedra. Valid inequalities for these sets are more easily related though. In particular, those valid for $\text{proj}_x(Z)$ are trivially valid for Z.

Proposition 2.2.10 *If $\pi^T x \leq \pi_0$ is valid for* $\text{proj}_x(Z)$*, then $\pi^T x \leq \pi_0$ is valid for Z.*

Proof Suppose that $\pi^T x \leq \pi_0$ is valid for $\text{proj}_x(Z)$. Let $(\bar{x}, \bar{y}) \in Z$. Then, $\bar{x} \in \text{proj}_x(Z)$ and so $\pi^T \bar{x} \leq \pi_0$. Hence, $(\bar{x}, \bar{y})$ satisfies the inequality $\pi^T x \leq \pi_0$, which is then valid for Z. ■

On the other hand, we can use Theorem 2.2.9 to derive valid inequalities for $\text{proj}_x(Z)$ from valid inequalities for Z, for example as follows.

Proposition 2.2.11 *If $\pi^T x + \lambda^T y \leq \pi_0$ is valid for $Z = \{(x,y) \in \mathbb{R}^p \times \mathbb{R}^q : Ax + By \leq b\}$, then $(\pi + \pi')^T x \leq \pi_0 + \pi'_0$ is valid for* $\text{proj}_x(Z)$ *for every $(\pi', \pi'_0) \in \mathbb{R}^p \times \mathbb{R}$ such that $U' := \{u \geq 0 : u^T A = (\pi')^T, u^T B = -\lambda^T\} \neq \emptyset$ and $\pi'_0 \geq \max\{u^T b : u \in U'\}$.*

Proof Suppose that $\pi^T x + \lambda^T y \leq \pi_0$ is valid for Z. Then, $Z = \{(x,y) \in \mathbb{R}^p \times \mathbb{R}^q : Ax + By \leq b, \pi^T x + \lambda^T y \leq \pi_0\}$. Let $(\pi', \pi'_0) \in \mathbb{R}^p \times \mathbb{R}$ such that $U' := \{u \geq 0 : u^T A = (\pi')^T, u^T B = -\lambda^T\} \neq \emptyset$ and $\pi'_0 \geq \max\{u^T b : u \in U'\}$. Let $\bar{u} \in U'$. Observe that $(\bar{u}, 1) \in \{(u,v) \geq 0 : u^T B + v\lambda^T = 0\}$. Using the equivalent expression of Z and Theorem 2.2.9, we conclude that $(\bar{u}^T A + \pi^T)x \leq \bar{u}^T b + \pi_0$ is valid for $\text{proj}_x(Z)$. Since $(\pi')^T = \bar{u}^T A$ and $\pi'_0 \geq \bar{u}^T b$, it follows that

$(\pi+\pi')^T x \leq \pi_0 + \pi'_0$ is valid for $\text{proj}_x(Z)$. ■

Corollary 2.2.12 *If $\pi^T x \leq \pi_0$ is valid for Z, then $\pi^T x \leq \pi_0$ is valid for* $\text{proj}_x(Z)$.

The dimension of $\text{proj}_x(Z)$ can be inferred from the dimension of Z ((BALAS; OOSTEN, 1998)). In particular, full-dimensional polyhedra are projected onto full-dimensional polyhedra.

Proposition 2.2.13 *If* $\dim(Z) = p+q$ *then* $\dim(\text{proj}_x(Z)) = p$.

Proof Suppose that $\dim(\text{proj}_x(Z)) < p$. Then, there is an equality $\pi^T x = \pi_0$ satisfied by every point $x \in \text{proj}_x(Z)$. Therefore, it is also satisfied by every point $(x,y) \in Z$, which implies that $\dim(Z) < p+q$. ■

As a particular case of Proposition 2.2.8, we have the following property.

Proposition 2.2.14 *Let $Z \subseteq \mathbb{R}^p \times \mathbb{R}^q$. Then,* $\text{proj}_x(\text{conv}(Z)) = \text{conv}(\text{proj}_x(Z))$.

Results and other concepts about polyhedra, linear programming, duality, *branch-and-bound* and *branch-and-cut* can be found in (FERREIRA; WAKABAYASHI, 1996), (SCHRIJVER, 1986), (NEMHAUSER; WOLSEY, 1988).

2.3 The set covering problem

Let $U = \{e_1, \ldots, e_m\}$ be a finite set of elements, $\mathscr{S} = \{S_1, \ldots, S_n\}$ be a given collection of subsets of U and $c = (c_1, \ldots, c_n)$ be a vector of costs, where $c_j \geq 0$, $\forall j = 1, \ldots, n$. Let F be an index subset of $\{1, \ldots, n\}$. F is said to *cover* U if $\bigcup_{j \in F} S_j = U$. The cost of F is $\sum_{j \in F} c_j$.

The *set covering problem* consists in determining a minimum-cost cover of U, if it exists. Note that $\bigcup_{j=1}^n S_j \neq U$ implies that there is no solution for the problem. The set covering problem can be stated as

$$(SC) \quad \min\{c^T x \mid Ax \geq 1,\ x \in \{0,1\}^n\},$$

where $A = (a_{ij})$ is an $m \times n$ matrix with $a_{ij} \in \{0,1\}$, $\forall i, j$, and $a_{ij} = 1$ if, and only if, $e_i \in S_j$. We use 1 as the m-vector of 1's. For a given general 0-1 matrix A, this problem is NP-hard ((KARP, 1972)).

The set covering polytope is defined as

$P_I(A) := conv\{x \in \mathbb{R}^n \mid Ax \geq 1,\ 0 \leq x \leq 1,\ x \text{ integer}\}.$

We also consider the polytope related to the linear relaxation of (SC):

$P(A) := conv\{x \in \mathbb{R}^n \mid Ax \geq 1,\ 0 \leq x \leq 1\}.$

Let M and N be the row and column index sets, respectively, of A. Assume that A has no zero rows or zero columns. From (BALAS; NG, 1989), we have the following statements about the set covering problem (SC):

1. $P_I(A)$ is full-dimensional if and only if $\sum_{j=1}^{n} a_{ij} \geq 2$ for all $i \in M$.

   In the following we assume that $P_I(A)$ is full-dimensional.
2. All facet defining inequalities $\alpha x \geq \alpha_0$ for $P_I(A)$ have $\alpha \geq 0$ if $\alpha_0 > 0$.
3. The inequality

   $$\sum_{j=1}^{n} a_{ij} x_j \geq 1$$

   defines a facet of $P_I(A)$ if and only if

   (F'1) there exists no $k \in M$ with $a_{kj} \leq a_{ij}$, $\forall j \in N$, and $\sum_{j=1}^{n} a_{kj} < \sum_{j=1}^{n} a_{ij}$;

   (F'2) for each k such that $a_{ik} = 0$, there exists $j(k)$ such that $a_{ij(k)} = 1$ and $a_{hj(k)} = 1$ for all $h \in M^0(k) := \{h \in M \mid a_{hk} = 1 \text{ and } a_{hj} = 0, \forall j \neq k, \text{ such that } a_{ij} = 0\}$.

4. The only minimal valid inequalities (hence the only facet-defining inequalities) for $P_I(A)$ with integer coefficients and right-hand side equal to 1 are those of the system $Ax \geq 1$.

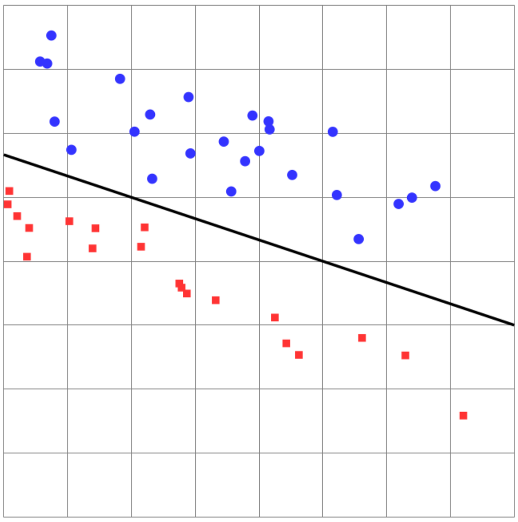
More results about valid inequalities and facet-defining properties for the set covering problem can be found in (CORNUÉJOLS; SASSANO, 1989) and (SÁNCHEZ-GARCÍA *et al.*, 1998).

3 EUCLIDEAN CLASSIFICATION PROBLEM

In this chapter, we formalize the Euclidean classification problem in a multidimensional space. Then, we show some solution methods based on mathematical programming found in the literature.

The elementary Euclidean classification problem in two classes can be described as follows. Let $\mathscr{S} = \{s_1, \ldots, s_m\}$ be a set of samples such that, for all $i \in [m]$, $[m] = \{1, \ldots, m\}$, we have $s_i = (x_i, y_i)$, $x_i \in \mathbb{R}^d$, $d \geq 1$ and $y_i \in \{1, 2\}$. The value of y_i indicates the class which point x_i belongs to. We use the notation $\mathscr{X} = \{x_1, \ldots, x_m\}$ for the set of points associated with $\mathscr{S}$, which is partitioned into $\mathscr{X}_1 = \{x_i \in \mathscr{X} \mid y_i = 1\}$ and $\mathscr{X}_2 = \{x_i \in \mathscr{X} \mid y_i = 2\}$. The general objective of the classification problem is a partitioning of the $\mathbb{R}^d$ space, based on the partition $\mathscr{X}_1$, $\mathscr{X}_2$ of the samples, and the assignment of each part to exactly one of the two classes. So, if $R \subseteq \mathbb{R}^d$ is the part assigned to class $y \in \{1, 2\}$, then every point $x \in R$ is classified as belonging to class y. The set associated with class y corresponds to the pattern established for this class.

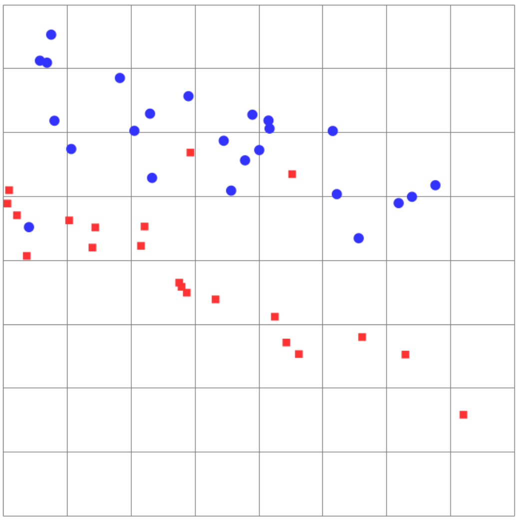
Figure 1 – Example of the Euclidean classification problem in $\mathbb{R}^2$. Circles are points of class 1, while squares represent points of class 2.



An example of instance for this classification problem is illustrated in Figure 1 where the marked line partitions the $\mathbb{R}^2$ space into two parts on the underlying pattern established by the points of each class. In this figure, it is shown the set of samples and a partitioning of the space. The situation in this example is such that the determination of the existing pattern in both classes is relatively simple. A bit more complex situation is illustrated in Figure 2 where there are class 1 points in the “middle"of class 2 points, and vice-versa. One sample $s_i \in S$ is an *outlier* (or a *discrepant* point) if y_i is inconsistent with the value of x_i according to the pattern established for class y_i. When sample s_i is an outlier, we also say that the corresponding point x_i is an outlier. An example of such a situation is shown in Figure 2. In this case, the partitioning presented in Figure 1 still seems to be a good solution. Its determination, however, requires the detection of the outliers and their disregard in the recognition of the patterns.

A generalization of this classification problem consists in considering multiple classes, that is, more than two classes. It is an interesting problem due to many associated applications. In this thesis, however, we only focus on the classification problem with two classes.

Figure 2 – Example of Figure 1 in which some points are misclassified, becoming discrepant points.



There are several solution methods for the Euclidean classification problem that are based on mathematical programming. Some of the most studied ones are summarized below. In this text, these methods are divided according to how the samples are separated to form the desired partition.

3.1 Linear separation

As in the example illustrated in Figure 1, the partitioning using linear separation is defined by a hyperplane $\mathbf{p}^T\mathbf{x}+q=0$, $\mathbf{p}\in\mathbb{R}^d$ and $q\in\mathbb{R}$. The classification of a point $\mathbf{x}'$ is made according to the Euclidean position in relation to that hyperplane. There are two possible cases: if $\mathbf{p}^T\mathbf{x}'+q<0$, then $\mathbf{x}'$ is classified in class 1; otherwise, class 2 is used to classify $\mathbf{x}'$. The existence of such a separating hyperplane, that correctly classifies every sample in S, is only ensured if this set has a special characteristic.

A set of S samples is *linearly separable* if the convex hulls of $\mathscr{X}_1$ and $\mathscr{X}_2$ are disjoint. In mathematical terms, a necessary and sufficient condition for S to be linearly separable is that there is no $\lambda = (\lambda_1, \ldots, \lambda_m)$ such that

$$\begin{aligned}
&\sum_{i \in [m], y_i = 1} \lambda_i x_i = \sum_{i \in [m], y_i = 2} \lambda_i x_i, \\
&\sum_{i \in [m], y_i = 1} \lambda_i = 1, \\
&\sum_{i \in [m], y_i = 2} \lambda_i = 1.
\end{aligned}$$

Then, Farkas's lemma allows us to conclude that a set of samples is linearly separable if, and only if, there are $\mathbf{p} \in \mathbb{R}^d$ and $q, r \in \mathbb{R}$ such that

$$\begin{aligned}
&r - q < 0 \\
&\mathbf{p}^T \mathbf{x}_i + q \leq 0, \ i \in [m], y_i = 1 \\
&\mathbf{p}^T \mathbf{x}_i + r \geq 0, \ i \in [m], y_i = 2.
\end{aligned} \tag{3.1}$$

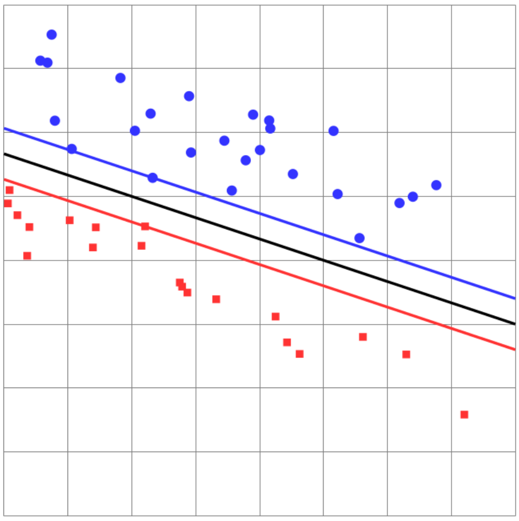
Since there is no constraint on the signal of q and r, we can define $\delta = q - r$ and then express condition (3.1) equivalently as the non-emptiness of any of the sets $(a) - (d)$ defined in Figure 3.

Figure 3 – Equivalent conditions for the existence of linear separation.

$r - q < 0$ $\mathbf{p}^\top \mathbf{x}_i + q \leq 0, i \in [m], y_i = 1$ $\mathbf{p}^\top \mathbf{x}_i + q \geq q - r, i \in [m], y_i = 2$ $\quad (a)$	$\mathbf{p}^\top \mathbf{x}_i + q \leq 0, i \in [m], y_i = 1$ $\mathbf{p}^\top \mathbf{x}_i + q \geq \delta, i \in [m], y_i = 2$ $\delta > 0$ $\quad (b)$
$\mathbf{p}^\top \mathbf{x}_i + q \leq -\delta, i \in [m], y_i = 1$ $\mathbf{p}^\top \mathbf{x}_i + q \geq \delta, i \in [m], y_i = 2$ $\delta > 0$ $\quad (c)$	$\mathbf{p}^\top \mathbf{x}_i + q \leq -1, i \in [m], y_i = 1$ $\mathbf{p}^\top \mathbf{x}_i + q \geq 1, i \in [m], y_i = 2$ $\quad (d)$

In particular, if S is linearly separable, then condition (c) in Figure 3 indicates that there exist $\mathbf{p}$, q and $\delta > 0$ such that $\mathbf{p}^T \mathbf{x}_i + q \leq -\delta$ is satisfied for all $x_i \in S$ with $y_i = 1$, and $\mathbf{p}^T \mathbf{x}_i + q \geq \delta$ is satisfied for all $x_i \in S$ with $y_i = 2$. Since $\mathbf{p}$ and q can be chosen such that $\delta = 1$, we get the condition (d) shown in Figure 3. Therefore, given p and q satisfying (d), the linear separation is given by the hyperplane $\mathbf{p}^T \mathbf{x} + q = 0$.

Figure 4 – Example of a separating hyperplane and margins of the classes for the example of Figure 1.



If a set of samples S is linearly separable, there may be several separating hyperplanes (satisfying (d)). In order to express the quality of a separator, the concept of margin is useful. A margin of a class is a hyperplane defining a face of the convex hull of the points in that class. If the set is linearly separable, there are parallel margins for the two classes. A good separator would lie halfway between them.

A well-known quadratic programming formulation aims at finding a linear separator satisfying (d) that maximizes the distance to the margins of the two classes.

Precisely, if the margins are defined by $\mathbf{p}^T\mathbf{x}+q=-\varepsilon$ and $\mathbf{p}^T\mathbf{x}+q=\varepsilon$, then the distance between them is equal to $\frac{2\varepsilon}{||\mathbf{p}||}$. Thus, maximizing such a distance is equivalent to minimizing $||\mathbf{p}||$ or still $\mathbf{p}^t\mathbf{p}$. This results in the formulation:

$$
\begin{aligned}
\min \quad & \frac{1}{2}\mathbf{p}^T\mathbf{p} \\
\text{s. t.} \quad & \mathbf{p}^T\mathbf{x}_i+q\leq -1,\; i\in[m], y_i=1, \\
& \mathbf{p}^T\mathbf{x}_i+q\geq 1,\; i\in[m], y_i=2, \\
& \mathbf{p}\in\mathbb{R}^d, \\
& q\in\mathbb{R}.
\end{aligned}
$$

A solution to this formulation, with the separating hyperplane and margins of the classes, is shown in Figure 4. Note that the separating hyperplane is defined by $\mathbf{p}^T\mathbf{x}+q=0$.

The above formulation leads to a solution method known as *SVM* (*Support Vectors Machine*), which is very efficient for linearly separable sample sets, according to several empirical studies.

3.2 Outlier points

Naturally, there are cases of sample sets that are not linearly separable, which occur very often in practice. One reason for that is the presence of outliers (see Figure 2). In this case, the classification problem can be defined as:

Problem 1. 2-*class Single-group Euclidean Classification Problem*:

Given the Euclidean space of dimension d, sets of initially classified samples $\mathscr{X}_1$ (blue points) and $\mathscr{X}_2$ (red points), find subsets $A_1 \subseteq \mathscr{X}_1$ and $A_2 \subset \mathscr{X}_2$ such that

$$conv(A_1) \cap conv(A_2) = \emptyset, \text{ and} \tag{3.2}$$

$$|(\mathscr{X}_1 \backslash A_1) \cup (\mathscr{X}_2 \backslash A_2)| \text{ is minimum.} \tag{3.3}$$

The subsets A_1 and A_2 consist of the non outlier points of each class. Thus, $(\mathscr{X}_1 \backslash A_1) \cup (\mathscr{X}_2 \backslash A_2)$ is the set of all outlier points. Condition (3.2) states that the sample set becomes linearly separable when disregarding the outliers. So, it ensures the existence of a separating hyperplane which leads to a classification of any initially non-classified point in $\mathbb{R}^d$, according to the half-space it belongs to. Besides, we want to find a solution with the minimum number of outliers, which is required by (3.3).

In such cases where outliers appear, the *SVM* method loses much of its efficiency, justifying the search for alternative formulations. One of these formulations is the following linear programming model derived by (BERTSIMAS; SHIODA, 2007):

$$\begin{array}{lll}
\min & \sum_{i \in [m]} \varepsilon_i & \\
\text{s. t.} & \mathbf{p}^T \mathbf{x}_i + q \leq -1 + \varepsilon_i, & i \in [m], y_i = 1, \\
& \mathbf{p}^T \mathbf{x}_i + q \geq 1 - \varepsilon_i, & i \in [m], y_i = 2, \\
& \varepsilon_i \geq 0, & i \in [m], \\
& \mathbf{p} \in \mathbb{R}^d, & \\
& q \in \mathbb{R}. &
\end{array}$$

Note that, instead of minimizing the number of outliers, the formulation minimizes the total violation (distance from an outlier to the half-space assigned with class y_i).

The formulation above separates the samples to minimize the violations ε_i, for all $i \in [m]$. If S is linearly separable, the optimal value of this problem is 0. Otherwise, the points $x_i \in \mathscr{X}$ with $\varepsilon_i > 0$ are considered outliers, and the remaining points are separated by the hyperplane defined by $\mathbf{p}$ and q.

A mixed integer programming formulation can be obtained from the previous linear formulation. It seeks to minimize the number of samples considered as outliers. Such a formulation, inspired by the formulation of (BERTSIMAS; SHIODA, 2007), uses the following variables:

- Binary variable: $o_i = 1$, if s_i is an outlier, and $o_i = 0$ otherwise.
- Real variables: $\mathbf{p}$ and q to define the separating hyperplane.
- Positive constant: M (big enough).

$$
\begin{array}{lll}
\min & \sum_{i\in[m]} o_i & \\
\text{s. t.} & \mathbf{p}^T\mathbf{x}_i + q \leq -1 + Mo_i, & i \in [m], y_i = 1, \\
& \mathbf{p}^T\mathbf{x}_i + q \geq 1 - Mo_i, & i \in [m], y_i = 2, \\
& o_i \in \{0,1\}, & i \in [m], \\
& \mathbf{p} \in \mathbb{R}^d, & \\
& q \in \mathbb{R}. &
\end{array}
$$

The formulation above is quite similar to the previous one. The difference is that the real variables ε_i are replaced by the binary variables o_i, for all $i \in [m]$. These new variables are used only to determine if an initially classified point (sample) is an outlier ($o_i = 1$) or not ($o_i = 0$). Thus, if s_i is considered as an outlier, the corresponding separator constraint must be ignored, which is assured by the big positive constant in the right-hand side.

Finally, a pure integer formulation is also found in the literature ((BLAUM *et al.*, 2019a)). It only uses variables $o_i \in \{0,1\}$, for all $i \in [m]$, to determine if s_i is an outlier or not [1].

[1] Actually, the original formulation in (BLAUM *et al.*, 2019a) uses variables $a_i = 1 - o_i, i \in [m]$, instead, to indicate whether x_i is not an outlier. We preferred to replace the variables to keep the correspondence with other formulations presented in this work.

It is

$$(\mathrm{ILP}_E)\min \sum_{i\in[m]} o_i$$

$$\text{s. t. } conv(\{s_i \mid o_i = 0, y_i = 1\}) \cap conv(\{s_i \mid o_i = 0, y_i = 2\}) = \emptyset, \tag{3.4}$$

$$o_i \in \mathbb{B}^m. \tag{3.5}$$

We denote by P_E the convex hull of the points satisfying constraints (3.4) and (3.5). It is worth noting that the linear separation constraint (3.4) in the formulation can be modeled by the inequalities

$$\sum_{x_i\in S\cup T} o_i \geq 1. \tag{3.6}$$

for every $S \subseteq \mathscr{X}_1$ and every $T \subseteq \mathscr{X}_2$ such that S and T are linearly inseparable, i.e. $conv(S) \cap conv(T) \neq \emptyset$. (BLAUM *et al.*, 2019a) studied the polytope associated with ILP_E.

3.2.1 $\mathscr{N}$-Set inequalities

In order to derive valid inequalities for ILP_E and study facetness properties, (BLAUM *et al.*, 2019a) introduced the following definition and notation. Let $(S \subseteq \mathscr{X}_1, T \subseteq \mathscr{X}_2)$ be linearly inseparable. An *$\mathscr{N}$-set* for (S,T) is a minimal $N \subseteq S \cup T$ such that $(S\backslash N, T\backslash N)$ is linearly separable. Let $\mathscr{N}(S,T) = \{N \subseteq S\cup T \mid N \text{ is an } \mathscr{N}\text{-set for } (S,T)\}$ and, for each $x_i \in S \cup T$,

$$\nu_i = \min\{|N| \mid N \in \mathscr{N}(S,T), x_i \in N\}.$$

We assume that $\nu_i = \infty$ if $\{N \in \mathscr{N}(S,T) \mid x_i \in N\} = \emptyset$ and $\frac{1}{\infty} = 0$. Also, we say that N is a *perfect $\mathscr{N}$-set* for (S,T) if $\nu_i = |N|$ for all $x_i \in N$. We define $\mathscr{N}^*(S,T) = \{N \mid N \text{ is a perfect } \mathscr{N}\text{-set for } (S,T)\}$.

Lemma 3.2.1 *Let $(S \subseteq \mathscr{X}_1, T \subseteq \mathscr{X}_2)$ be linearly inseparable. The following $\mathscr{N}$-set inequality is valid for P_E:*

$$\sum_{x_i\in S\cup T} \frac{o_i}{\nu_i} \geq 1. \tag{3.7}$$

Proof Let $\bar{o}$ be a feasible solution for ILP_E and $N' = \{x_i \in S\cup T \mid \bar{o}_i = 1\}$. Since $\bar{o}$ is a feasible solution, $N' \neq \emptyset$. Then, there exists $N \subseteq N'$ such that $N \in \mathscr{N}(S,T)$, which leads to $\sum_{x_i\in S\cup T} \frac{\bar{o}_i}{\nu_i} = \sum_{x_i\in N'} \frac{1}{\nu_i} \geq \sum_{x_i\in N} \frac{1}{\nu_i} \geq \sum_{x_i\in N} \frac{1}{|N|} = 1$, as $\nu_i \leq |N|, \forall x_i \in N$. ■

Figure 5 – Examples of $\mathcal{N}$-set inequalities that define facets of P_E ((BLAUM *et al.*, 2019a)).

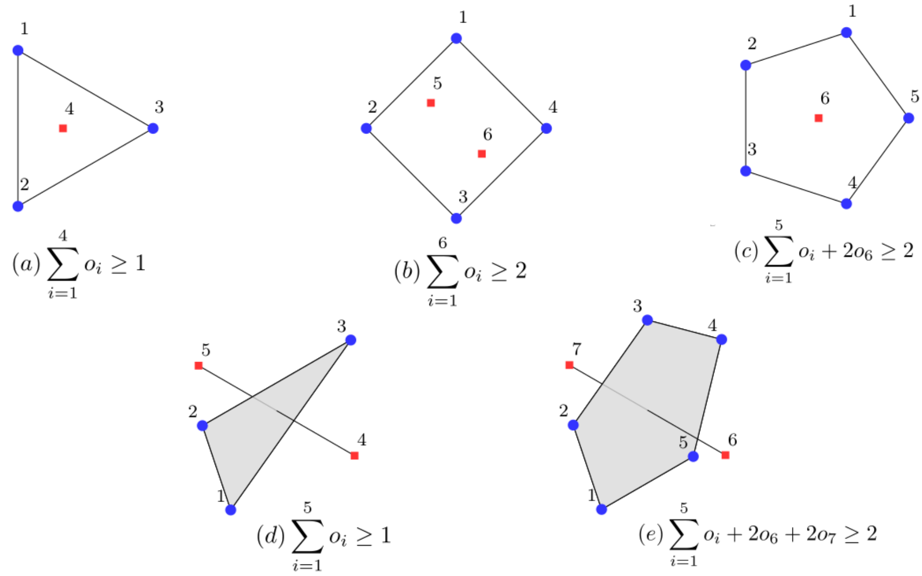


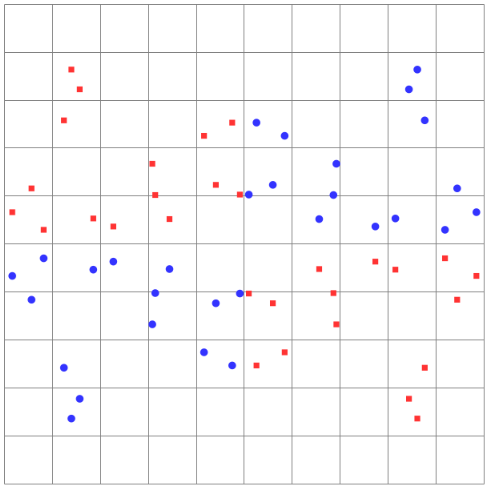
Figure 5 shows some examples of such inequalities. In each example, the depicted points define a linearly inseparable (S,T) pair. By the theorem below, these examples define facets of P_E.

Theorem 3.2.2 ((CORRÊA *et al.*, 2019)) *The inequality* (3.7) *defines a facet of* P_E *if* ($|T| = 1$ *or* ($|T| = 2$ *and* $T \cap conv(S) = \emptyset$)) *and* S *is minimal with respect to the property* $conv(T) \cap conv(S) \neq \emptyset$ *(i.e.,* $conv(T) \cap conv(S') = \emptyset$*, for every* $S' \subset S$*).*

It is worth observing that the formulations presented in this subsection assume that the reason for the samples set to be linearly inseparable is solely the presence of outliers. However, there are cases in which the linear inseparability is due to the own nature of the samples set (see Figure 6). A way of dealing with such a situation is to transform the samples $s_i = (x_i, y_i) \in S$ to samples $(g(\mathbf{x}_i), y_i)$, where g is a function that maps a point $x_i \in \mathbb{R}^d$ to a point in an Euclidean space of dimension d' greater than d. The transformation aims to turn the set of transformed samples linearly separable. Thus, finding a separating hyperplane in such a higher dimension space corresponds to a non-linear separation in the original space $\mathbb{R}^d$. This strategy, when used with *SVM*, may lead to an efficient classification method when the set of samples does not contain discrepant points. However, one may need to consider a very high dimension d' when the samples are highly linearly inseparable.

Another way to deal with non-linear separability is discussed next.

Figure 6 – An instance for the classification problem in which the sample set is linearly inseparable and poorly classified with a single separating hyperplane.



3.3 Piecewise linear separation

We now show a *piecewise linear separation* approach for classification and present two integer linear programming models. This separation approach is best suited to cases where the set of samples is linearly inseparable due to its own nature. To deal with this scenario, the idea is to partition $\mathscr{X}_1$ and $\mathscr{X}_2$ into subsets, called *groups*, for better identification of the underlying patterns, while simultaneously eliminating discrepant points. As a result, we have the families $\{\mathscr{X}_1^1, \ldots, \mathscr{X}_1^{L_1}\}$ and $\{\mathscr{X}_2^1, \ldots, \mathscr{X}_2^{L_2}\}$, where L_1 and L_2 are constants previously defined. A constraint is imposed for the grouping, namely: after the removal of the discovered discrepant points, the subset of points in $\mathscr{X}_1^k \cup \mathscr{X}_2^l$, $k \in \{1, \ldots, L_1\} = [L_1]$ and $l \in \{1, \ldots, L_2\} = [L_2]$, must be linearly separable (see Figure 7). The linear separation of each pair $(\mathscr{X}_1^k, \mathscr{X}_2^l)$ generates a separating hyperplane $\mathbf{p}_{k,l}^T \mathbf{x} + q_{k,l} = 0$, which separates the space into two half-spaces, leaving the points of $\mathscr{X}_1^k$ within one of them and the points of $\mathscr{X}_2^l$ within the other one. For each $k \in [L_1]$, the polyhedron of group k is given by the intersection of the half-spaces defined by the hyperplanes separating $\mathscr{X}_1^k$ and $\mathscr{X}_2^l$, for each $l \in [L_2]$. The union of the polyhedra of all groups in $[L_1]$ form the points of the space assigned to class 1. The points assigned to class 2 are those of the polyhedra of groups $l \in [L_2]$, similarly defined. Notice that there may exist a point not covered by the union of these polyhedra. The class of any such a point is arbitrarily chosen (see Figure 7).

This classification problem can be defined as below:

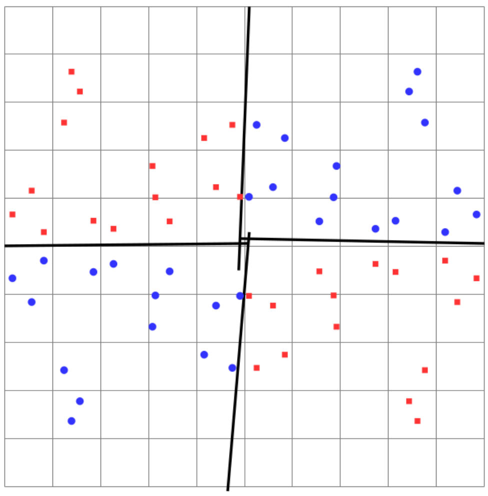
Problem 2. 2-*class Multi-group Euclidean Classification Problem*:

Given the Euclidean space of dimension d, sets of initially classified samples $\mathscr{X}_1$ (blue points) and $\mathscr{X}_2$ (red points), and upper bounding parameters L_1, L_2, find groups $A_k \subseteq \mathscr{X}_1$, $k \in [L_1]$, and $A_{k'} \subset \mathscr{X}_2$, $k' \in [L_2]$, such that

1. $conv(A_k) \cap conv(A_{k'}) = \emptyset$, for all $k \in [L_1]$, $k' \in [L_2]$, and
2. $|\mathscr{X}_1 \cup \mathscr{X}_2| - |(\bigcup_k A_k \cup \bigcup_{k'} A_{k'})|$ is minimum (which is the number of outlier points).

For any $k \in [L_1]$, $conv(A_k)$ is called a blue (class 1) convex set. Similarly, we name $conv(A_{k'})$ the red (class 2) convex sets. Thus, points in $(\mathbb{R}^d \backslash (\mathscr{X}_1 \cup \mathscr{X}_2))$ that belong to a blue (resp., red) convex set are set to the blue (resp., red) class. If a point $x_i \in \mathscr{X}_1$ (resp., $x_i \in \mathscr{X}_2$) does not belong to $\bigcup_k A_k$ (resp., $\bigcup_{k'} A_{k'}$), then x_i is an outlier. It may happen that a point in $(\mathbb{R}^d \backslash (\mathscr{X}_1 \cup \mathscr{X}_2))$ does not belong to any colored convex set. In such cases, we arbitrarily choose a class for it. An example of a solution in such a form is illustrated in Figure 7.

Figure 7 – Hyperplanes to separate each pair of groups of opposite classes for the example of Figure 6, with $L_1 = L_2 = 2$.



A general mathematical formulation for this approach is given by the model described in (BERTSIMAS; SHIODA, 2007) and (SHIODA, 2003). Its associated polyhedron is studied in (CORRÊA *et al.*, 2019). The variables for this formulation are:

- Binary variables: $a_{ki} = 1$, if $x_i \in \mathscr{X}_{y_i}^k$, and $a_{ki} = 0$ otherwise, for all $i \in [m]$ and $k \in [L_{y_i}]$. The role of this variable is to activate the corresponding separation constraint in the formulations below only if point x_i is assigned to group k.
- Real variables: $\varepsilon_i \geq 0$. Used to indicate how much a point $x_i \in \mathscr{X}$ violates any separating hyperplane constraint, if it is the case.

- Real variables: $\mathbf{p}_{kl}^T$ and q_{kl} to define the separating hyperplanes for each $k \in [L_1]$ and $l \in [L_2]$.
- Positive constant: M (big enough).

With these elements, two formulations arise:

- Formulation minimizing outlier violations:

$$
\begin{array}{lrl}
\min & \sum_{i\in[m]} \varepsilon_i & \\
\text{s. t.} & \mathbf{p}_{kl}^T\mathbf{x}_i + q_{kl} \le M-(M+1)a_{ki}+\varepsilon_i, & i\in[m], y_i=1, k\in[L_1], l\in[L_2],\\
& \mathbf{p}_{kl}^T\mathbf{x}_i + q_{kl} \ge -M+(M+1)a_{li}-\varepsilon_i, & i\in[m], y_i=2, k\in[L_1], l\in[L_2],\\
& \sum_{k\in[L_{y_i}]} a_{ki} = 1, & i\in[m],\\
& a_{ki}\in\{0,1\}, & i\in[m], k\in[L_{y_i}],\\
& \varepsilon_i \ge 0, & i\in[m],\\
& \mathbf{p}_{kl}\in\mathbb{R}^d, & k\in[L_1], l\in[L_2],\\
& q_{kl}\in\mathbb{R}, & k\in[L_1], l\in[L_2].
\end{array}
$$

- Formulation minimizing the amount of outliers:

$$
\begin{array}{lrl}
\max & \sum_{i\in[m]}\sum_{k\in[L_{y_i}]} a_{ki} & \\
\text{s. t.} & \mathbf{p}_{kl}^T\mathbf{x}_i + q_{kl} \le M-(M+1)a_{ki}, & i\in[m], y_i=1, k\in[L_1], l\in[L_2],\\
& \mathbf{p}_{kl}^T\mathbf{x}_i + q_{kl} \ge -M+(M+1)a_{li}, & i\in[m], y_i=2, k\in[L_1], l\in[L_2],\\
& \sum_{k\in[L_{y_i}]} a_{ki} \le 1, & i\in[m],\\
& a_{ki}\in\{0,1\}, & i\in[m], k\in[L_{y_i}],\\
& \mathbf{p}_{kl}\in\mathbb{R}^d, & k\in[L_1], l\in[L_2],\\
& q_{kl}\in\mathbb{R}, & k\in[L_1], l\in[L_2].
\end{array}
\tag{3.8}
$$

Each of the formulations above defines one separating hyperplane for each pair of groups of opposite classes, precisely $\mathbf{p}_{kl}^T\mathbf{x} + q_{kl} = 0$, for each $k \in [L_1]$ and $l \in [L_2]$. The difference between the formulations is that the first one determines the outlier points using the ε variables ($\varepsilon_i > 0$ if, and only if, $x_i \in \mathscr{S}$ is an outlier) while the second one determines them by the variables a ($\sum_{k\in[L_{y_i}]} a_{ki} = 0$ means x_i is an outlier).

These formulations include a relatively large amount of binary variables. In (BERTSIMAS; SHIODA, 2007), a strategy is presented to reduce this number of variables through a prior step of *clustering*, and then a later grouping of the clusters thus obtained. In (CORRÊA *et al.*, 2019), the authors explore the integer programming aspects of the classification part of *CRIO* ((BERTSIMAS; SHIODA, 2007)). They deduce facet-inducing inequalities coming from the stable set polytope, showing that this classification problem has exploitable combinatorial properties.

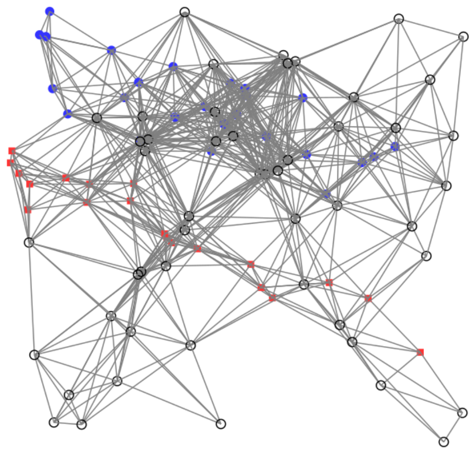
In the next chapter, we define a new classification problem, called geodesic classification problem, which has similarities between samples defined by non-numerical relations. The solution method for such a new problem was inspired by formulation ILP_E, as we can see in the integer formulation models also in the next chapter. Thus, we focus on minimizing the number of outliers as the objective criterion of the problem.

4 GEODESIC CLASSIFICATION PROBLEM

In this chapter, we define the notions of linear separation and classification on graphs using the geodesic convexity. We define convexity constraints to be used to classify data on graphs and study the associated polyhedron.

When we use a graph to model similarities, the vertices represent the elements of a universe while an edge between two vertices indicates some notion of similarity between them. Thus, by making a parallel between the graph and the Euclidean counterpart, we can use the former as input for a classification problem. Precisely, we are given a connected graph G where two disjoint subsets $V_B, V_R \subseteq V(G)$ are identified, respectively, as belonging to the *blue* and *red* classes. These are called *initially classified* vertices and correspond to the samples. The remaining ones are the unclassified (*neutral*) vertices and define the set V_N. These vertices are the analog of the points in the space different from the samples. An example of such a graph, to be called *classification graph*, is illustrated in Figure 8.

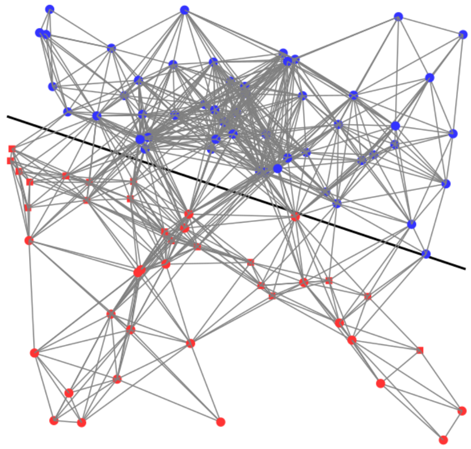
Figure 8 – An example of a classification graph. The filled vertices are those associated with the two classes.



The goal is to classify the vertices in V_N as blue or red by partitioning the vertices of the classification graph according to some convexity pattern (e.g., geodesic convexity). See Figure 9.

In the classification context, a mapping from points in the space to a classification graph can be obtained by defining an edge for each pair of points that are within a certain threshold distance. However, other similarity metrics can be applied. The classification graph can model classification scenarios where the objects are not numerically encoded, and similarities are measured by taking into account qualitative information. Most prominent examples are

Figure 9 – Vertex partition of a classification graph according to the geodesic convexity.



complex networks, such as social networks, networks of citations of scientific articles, etc. An analogy between the convexity concepts in discrete and continuous mathematics can be established if we consider the vertex set of a connected graph and the distance between vertices as a metric space. Thus, a vertex v is a convex combination of two other vertices u and w if v belongs to a shortest path between u and w (geodesic convexity). Other definitions of convexity on graphs have already been studied.

In analogy to the Euclidean version of the classification problem, we introduce the following definition.

Definition 4.0.1 *A triple* $(A_B \subseteq V_B, A_R \subseteq V_R, A_N \subseteq V_N)$ *is* linearly separable (with respect to G) *if*

(C1) $H[A_B] \cap A_R = \emptyset$,

(C2) $H[A_R] \cap A_B = \emptyset$, *and*

(C3) $H[A_B] \cap H[A_R] \cap A_N = \emptyset$,

and linearly inseparable *otherwise. For the sake of simplicity, we refer to* (A_B, A_R, A_N) *as* (A_B, A_R) *if* $A_N = V_N$.

Recall that the convex hull $H[S]$ of $S \subseteq V$ is the minimum convex set (related to the geodesic convexity) containing S. The subsets A_B and A_R are called the *basis* of the blue and red classes, respectively. Each basis *spans* on the graph through an operator to express the pattern of the corresponding class. In this case, the operator is given by the convex hull.

Figure 10 presents a classification graph where the filled circles represent the vertices in V_B (blue vertices) and the squares represent the vertices in V_R (red vertices). In this example, (V_B, V_R) is linearly inseparable. Indeed, v is simultaneously in a shortest path between two red vertices (v_1, v_2) and two blue vertices (v_3, v_4). Thus, $v \in H[V_B] \cap H[V_R]$ so that condition (C3) is violated for $(A_B, A_R) = (V_B, V_R)$. However, $(V_B \setminus \{v_i\}, V_R)$, for any $i \in \{3,4\}$, as well as $(V_B, V_R \setminus \{v_i\})$, for any $i \in \{1,2\}$, are linearly separable.

As a parallel with the Euclidean case, we say that (V_B, V_R) “becomes"linearly separable when considering any of v_1, v_2, v_3, v_4 as an outlier. This separation leads to the classification of the vertices in V_N. For instance, if we consider v_4 as an outlier, we get the classification depicted in Figure 11. It is worth remarking that considering a vertex $w \in V_B \cup V_R$ as an outlier does not mean removing it from the graph. It only signifies that w is considered neither red nor blue when calculating the convex hull of the red vertices or blue vertices.

Figure 10 – An example of a classification graph where (V_B, V_R) is linearly inseparable and $(V_B \setminus \{v_4\}, V_R)$ is linearly separable. Filled circles are vertices of V_B, while squares are vertices of V_R.

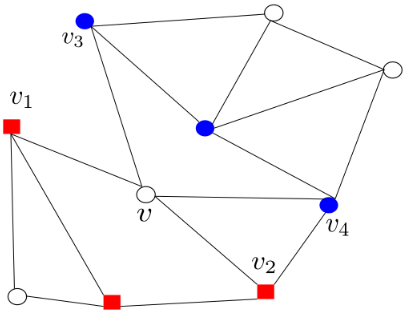
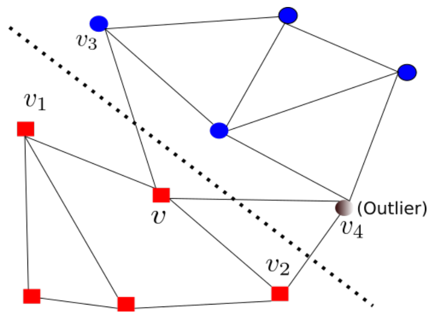


Figure 11 – An example of solution with an outlier for the example of Figure 10.



In terms of the linear separability defined above, we could say that (A_B, A_R) is linearly separable if, and only if, (V_B, V_R) becomes linearly separable when $V_B \setminus A_B$ and $V_R \setminus A_R$ are taken as outliers. Consequently, the class of any unclassified vertex can be established. With this in mind, we can introduce the geodesic classification problem associated with G, V_B and V_R. It consists of determining the smallest number of outlier vertices in order to make (V_B, V_R) linearly separable. More formally, we define:

Problem 3. 2-*class Single-group Geodesic Classification Problem* (2-*SGC*):

Given a connected graph $G = (V, E)$, sets of initially classified vertices V_B (blue vertices) and V_R (red vertices), and $V_N = V \backslash (V_{BR})$, find subsets $A_B \subseteq V_B$, $A_R \subseteq V_R$ such that:

(C0) (A_B, A_R) satisfies (C1), (C2) and (C3), and

(C4) $|V_{BR}| - |A_B \cup A_R|$ is minimum.

The vertices in $V_{BR} \setminus (A_B \cup A_R)$ are the outliers. Conditions (C1) and (C2) ensure that if an initially classified vertex i belongs to the convex hull of the non-outlier vertices of its opposite class, then i must be an outlier, i.e., $i \notin A_B \cup A_R$. Since every initially unclassified vertex needs to be assigned to exactly one class, it must belong to at most one convex hull of non-outliers of the same class. This is guaranteed by Condition (C3). Moreover, we want to find a solution with the minimum number of outliers, which is required by Condition (C4).

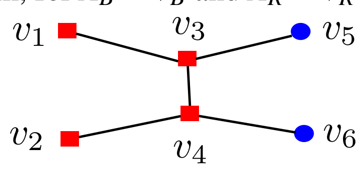
It is worth observing that Problem 2-SGC always has a solution. In the worst case, we could consider all initially classified vertices of a class as outliers. Besides, note that we can define this classification problem using other objective functions. However, to implement an application of the geodesic convexity approach for computational experiments, we use the same objective function (minimization of the number of outliers) as in ILP_E (see Section 3.2).

Any feasible solution (A_B, A_R) of Problem 2-SGC defines a mapping from V onto $\{blue, red\}$, which classifies all neutral vertices in $H[A_B]$ as blue class ones and all neutral vertices in $H[A_R]$ as red class ones. Note that there may exist vertices in $V_N \cap (V \setminus (H[A_B] \cup H[A_R]))$. However, the classification of such vertices is not in the scope of this work. Due to the outliers and the vertices in $V_N \cap (V \setminus (H[A_B] \cup H[A_R]))$, each class does not necessarily define a convex set. Besides, $(H[A_B], H[A_R])$ is neither a covering nor a packing of the vertices of G.

Observe that the definition of linear separability for the geodesic classification (Conditions (C1)-(C3)) is different from the Euclidean counterpart (Constraint (3.2)). It is because the relaxed conditions (C1)-(C3) allow to have more feasible solutions that are more likely to appear in the geodesic version of the classification problem. As an example, consider the picture in Figure 12. If we define the linear separability in the geodesic version as $H[A_B] \cap$

$H[A_R] = \emptyset$, instead of (C1)-(C3), then we could not have $(A_B = \{v_5, v_6\}, A_R \subseteq \{v_1, v_2, v_3, v_4\})$, with $|A_R| > 1$, as a feasible solution, since $H[A_B] \cap H[A_R] \neq \emptyset$.

Figure 12 – An example where $H[A_B] \cap H[A_R] \neq \emptyset$ but $H[A_B] \cap H[A_R] \cap V_N = \emptyset$ in the geodesic classification problem, for $A_B = V_B$ and $A_R = V_R \setminus \{v_3, v_4\}$.



4.1 Integer formulations for the 2-SGC problem

We present two integer linear formulations for the 2-SGC Problem and show that one is a projection of the other. The first formulation has a linear number of variables but an exponential number of constraints, whereas the second one has more variables but a polynomial number of constraints. An interesting feature of the first model is that it expresses the 2-SGC problem as a set covering problem. Thus, we can take advantage of well-known results from the literature.

4.1.1 A set covering formulation for the 2-SGC problem

For the first integer linear formulation, we define a binary variable o_i for each vertex $i \in V_{BR}$. A vertex $i \in V_{BR}$ is an outlier if, and only if, $o_i = 1$. Thus, we have

$$o_i = \begin{cases} 1, & \text{if } i \in V_{BR} \setminus (A_B \cup A_R), \\ 0, & \text{if } i \in (A_B \cup A_R). \end{cases}$$

For every $i \in V_{BR}$, let $K(i) \in \{B, R\}$ be the class of vertex i and $\bar{K}(i)$ be the opposite class to $K(i)$. Besides, if $K \in \{B, R\}$, then $\bar{K}$ is the opposite class to K. Then, the following set covering formulation is valid for the 2-SGC Problem.

$$\text{(ILP1)} \ \min \sum_{i \in V_{BR}} o_i$$

$$\text{st:} \sum_{j \in S \cup \{i\}} o_j \geq 1, \qquad \forall i \in V_{BR}, \forall S \subseteq V_{\bar{K}(i)} : i \in H[S], \tag{4.1}$$

$$\sum_{j \in S \cup T} o_j \geq 1, \qquad \forall S \subseteq V_B, \forall T \subseteq V_R : H[S] \cap H[T] \cap V_N \neq \emptyset, \tag{4.2}$$

$$o \in \mathbb{B}^{|V_{BR}|}. \tag{4.3}$$

Proposition 4.1.1 *Formulation ILP*1 *is correct.*

Proof Let $o \in \mathbb{B}^{|V_{BR}|}$ be a feasible solution of ILP1. Define $A_K = \{i \in V_K : o_i = 0\}$, for $K \in \{B, R\}$. Suppose that condition (C1) is not satisfied. Then, there exists $i \in A_R \cap H[A_B]$. Let $S = A_B \subseteq V_{\bar{K}(i)}$. Then, $o_j = 0$ for all $j \in S \cup \{i\}$, which violates a constraint in (4.1): a contradiction. Similarly, we show that condition (C2) holds. Now, suppose that condition (C3) is violated. Thus, constraint (4.2) is violated for $S = A_B$ and $T = A_R$: a contradiction. Therefore, (C1)-(C3) are satisfied.

Now, let $A_B \subseteq V_B$ and $A_R \subseteq V_R$ satisfying (C1)-(C3). Define $o \in \mathbb{B}^{|V_{BR}|}$ such that $o_i = 0$ if, and only if, $i \in A_B \cup A_R$. We first consider constraints (4.1). Let $i \in V_R$ and $S \subseteq V_B$ such that $i \in H[S]$. If $S \subseteq A_B$, by condition (C1) we must have $i \in V_R \setminus A_R$, and so $o_i = 1$. Otherwise, there is $j \in S$ such that $j \in V_B \setminus A_B$, and so $o_j = 1$. In both cases, constraint (4.1) is satisfied. Using condition (C2), we get a similar result for $i \in V_B$ and $S \subseteq V_R$. To consider constraints (4.2), let $S \subseteq V_B$ and $T \subseteq V_R$ such that $H[S] \cap H[T] \cap V_N \neq \emptyset$. By condition (C3), $S \setminus A_B \neq \emptyset$ or $T \setminus A_R \neq \emptyset$. Therefore, there is $j \in S \cup T$ such that $o_j = 1$, showing that constraint (4.2) is satisfied.

By the definition of the variables, the objective function trivially sums the number of vertices in $V_{BR} \setminus (A_B \cup A_R)$, which has to be minimized. ■

Consider the following subset of constraints of ILP1:

$$\sum_{j \in S \cup \{i\}} o_j \geq 1, \qquad \forall i \in V_{BR}, \forall S \subseteq V_{\bar{K}(i)} : i \in H[S], S \text{ minimal}, \tag{4.4}$$

$$\sum_{j \in S \cup T} o_j \geq 1, \qquad \forall S \subseteq V_B, \forall T \subseteq V_R : H[S] \cap H[T] \cap V_N \neq \emptyset, \tag{4.5}$$

$$H[S] \cap T = H[T] \cap S = \emptyset, S \cup T \text{ minimal},$$

where, in (4.4), S is minimal with regard to $i \in H[S]$ (i.e, $i \notin H[S \setminus \{u\}]$, for every $u \in S$) and, in (4.5), $S \cup T$ is minimal with regard to $H[S] \cap H[T] \cap V_N \neq \emptyset$ (i.e, $H[S \setminus \{u\}] \cap H[T] \cap V_N = \emptyset$, for every $u \in S$, and $H[S] \cap H[T \setminus \{v\}] \cap V_N = \emptyset$, for every $v \in T$).

We show that every constraint in (4.1)-(4.2) not belonging to (4.4)-(4.5) is redundant.

Proposition 4.1.2 *Constraints* (4.1) *and* (4.2) *are dominated by* (4.4) *and* (4.5).

Proof Let $K \in \{B,R\}$, $S \subseteq V_K$ and $i \in V_{\bar{K}}$ such that S is not minimal with regard to $i \in H[S]$. Then, there is $N \subset S$, $N \neq \emptyset$, such that $i \in H[S \backslash N]$ and $S \setminus N$ is minimal. So, $\sum_{j \in (S \backslash N) \cup \{i\}} o_j \geq 1$ is valid for ILP1 and dominates $\sum_{j \in S \cup \{i\}} o_j \geq 1$. Therefore, (4.1) is dominated by (4.4).

Now, let $S \subseteq V_B$ and $T \subseteq V_R$ such that $H[S] \cap T = H[T] \cap S = \emptyset$ is not true or $S \cup T$ is not minimal with regard to $H[S] \cap H[T] \cap V_N \neq \emptyset$. First, observe that, if $H[S] \cap T = H[T] \cap S = \emptyset$ is not true, then $H[S] \cap T \neq \emptyset$ or $H[T] \cap S \neq \emptyset$. In both cases, (4.2) is dominated by (4.1) and, consequently, by (4.4). If $S \cup T$ is not minimal with regard to $H[S] \cap H[T] \cap V_N \neq \emptyset$, then there is $N \subset S \cup T$, $N \neq \emptyset$, such that $H[S \backslash N] \cap H[T \backslash N] \cap V_N \neq \emptyset$ and $(S \cup T \setminus N)$ is minimal. So, $\sum_{j \in S \cup T} o_j \geq 1$ is dominated by $\sum_{j \in (S \cup T) \backslash N} o_j \geq 1$. It implies that (4.2) is dominated by (4.5). ■

4.1.2 ILP1's polyhedron

In this section, we study the polyhedron associated with the formulation ILP1. Since ILP1 forms a set covering problem, some of the proofs can be obtained as corollaries of the theorems in (BALAS; NG, 1989). However, in these cases, we present an alternative proof because parts of them are used in other following proofs.

Let $P_1 = conv(\{o \in \mathbb{B}^{|V_{BR}|} \mid o \text{ satisfies (4.1) and (4.2)}\})$ and $|V_{BR}| = n_1$. Next, we prove two basic properties of P_1.

Proposition 4.1.3 *P_1 is a full-dimensional polyhedron.*

Proof Consider the following points in $\mathbb{R}^{n_1}$: $f^0 = e$ and $f^1, \ldots, f^{n_1}$, where, for $i \in \{1, \ldots, n_1\}$, $f^i = e - e^i$. Clearly, these are $n_1 + 1$ affinely independent points. To show that they are feasible solutions of ILP1, first note that point f^0 has $f_i^0 = 1, \forall i \in V_{BR}$, and so it satisfies constraints (4.1) and (4.2). Let $i \in V_{BR}$. For point f^i, there is only one vertex u such that $f_u^i = 0$, which is $u = i$. Observe that, for every constraint in (4.1), S is a non-empty set and $i \notin S$. Then, there is at least one vertex $j \in S$ such that $j \neq i$, and so $f_j^i = 1$, which implies that f^i satisfies (4.1). Since every constraint in (4.2) involve two disjoint and non-empty subsets S and T, there is at least one vertex $j \in S \cup T$ such that $j \neq i$, and so $f_j^i = 1$, which yields that f^i satisfies (4.2).

Therefore, P_1 is a full-dimensional polyhedron. ■

Proposition 4.1.4 *Let $\pi^T o \geq \pi_0$ be a facet-defining inequality of P_1. If it is different from $o_i \leq 1$ and $o_i \geq 0$, for all $i \in V_{BR}$, then $\pi \geq 0$ and $\pi_0 > 0$.*

Proof Let $i \in V_{BR}$. Since the inequality is different from $o_i \leq 1$, the facet $F := \{o \in P_1 : \pi^T o = \pi_0\}$ contains a point $\bar{o}$ with $\bar{o}_i = 0$. Besides, the point o' such that $o'_j = \bar{o}_j$, for all $j \neq i$, and $o'_i = 1$ belongs to P_1. Therefore, $\pi^T(o' - \bar{o}) \geq \pi_0 - \pi_0 = 0$, which leads to $\pi_i \geq 0$. Since $\pi \neq 0$, there must be $i \in V_{BR}$ such that $\pi_i > 0$. As before, since the inequality is different from $o_i \geq 0$, it contains a point $\hat{o}$ with $\hat{o}_i = 1$. Then, $\pi_0 = \pi^T \hat{o} \geq \pi_i > 0$. ■

4.2 $\mathcal{N}$-set inequalities

Similarly to the Euclidean version, we use the following definitions. Let $(S \subseteq V_B, T \subseteq V_R)$ be linearly inseparable according to Definition 4.0.1. An *$\mathcal{N}$-set* for (S,T) is a minimal $N \subseteq S \cup T$ such that $(S \backslash N, T \backslash N)$ is linearly separable. We define $\mathcal{N}(S,T) = \{N \subseteq S \cup T \mid N$ is an $\mathcal{N}$-set for $(S,T)\}$, and for each $i \in S \cup T$,

$$\nu_i = \min\{|N| \mid N \in \mathcal{N}(S,T), i \in N\}.$$

We assume that $\nu_i = \infty$ if $\{N \in \mathcal{N}(S,T) \mid i \in N\} = \emptyset$. Also, we say that N is a *perfect $\mathcal{N}$-set* for (S,T) if $\nu_i = |N|$ for all $i \in N$. We define $\mathcal{N}^*(S,T) = \{N \mid N$ is a perfect $\mathcal{N}$-set for (S,T)$\}$. Observe that the concepts above are the same introduced by (BLAUM *et al.*, 2019a) for the Euclidean case, but now using the notion of linear separability given in Definition 4.0.1.

Proposition 4.2.1 *Let $(S \subseteq V_B, T \subseteq V_R)$ be linearly inseparable. The following $\mathcal{N}$-set inequality is valid for P_1:*

$$\sum_{i \in S \cup T} \frac{o_i}{\nu_i} \geq 1. \tag{4.6}$$

Proof Same proof of Proposition 3.2.1. ■

In the expression of inequality (4.6) we can assume, without loss of generality, that $\nu_i < \infty$ for all $i \in S \cup T$. Otherwise, $(S \setminus I, T \setminus I)$, for $I = \{i \in S \cup T \mid \nu_i = \infty\}$, is also linearly inseparable and defines the same inequality. If $(S \setminus I, T \setminus I)$ were linearly separable, there would be an $\mathcal{N}$-set $N \subseteq I \subseteq (S \cup T)$ and we would have $\nu_i < \infty$ for all $i \in N \subseteq I$.

For each $k \in \mathbb{N}$, the *safe graph* $G^k_{S,T} = (V^k_{S,T}, E^k_{S,T})$ is defined by $V^k_{S,T} = \{i \in S \cup T \mid \nu_i = k\}$ and $E^k_{S,T} = \{ij \mid \exists N_i, N_j \in \mathcal{N}^*(S,T), i \in N_i, j \in N_j, N_i \triangle N_j = \{i,j\}\}$. The union of all such graphs is $G_{S,T} = (V_{S,T}, E_{S,T})$ with $V_{S,T} = \bigcup_{k \in \mathbb{N}} V^k_{S,T}$ and $E_{S,T} = \bigcup_{k \in \mathbb{N}} E^k_{S,T}$.

This notion of safe graph was introduced by (CAMPÊLO *et al.*, 2008) for the study of the coloring polytope associated with the asymmetric representatives formulation. It is very useful to prove general facetness conditions. Notice that the coloring problem can be seen as a covering problem by independent sets. This notion was also used by (BLAUM *et al.*, 2019a) for their polyhedral studies associated with the Euclidean classification problem, where Theorem 4.2.2, shown below, was introduced. This theorem shows sufficient conditions for inequalities (4.6) to be facet-defining. Its proof is strongly inspired by the one in (BLAUM *et al.*, 2019b).

Theorem 4.2.2 *Let* $(S \subseteq V_B, T \subseteq V_R)$ *be linearly inseparable. Then,* (4.6) *defines a facet of* P_1 *if all the following assertions hold:*

(F1) for each $k > 1$, $V^k_{S,T} = \emptyset$ *or* ($|V^k_{S,T}| > 1$ *and* $G^k_{S,T}$ *is connected), and*

(F2) for each $i \in V_B \backslash S$ *(resp.* $j \in V_R \backslash T$*), there exists an* $N \in \mathscr{N}^*(S,T)$ *such that* $(S \cup \{i\} \backslash N, T \backslash N)$ *(resp.* $(S \backslash N, T \cup \{j\} \backslash N)$*) is linearly separable.*

Proof Let F be the face of P_1 defined by (4.6), and suppose $\lambda^T o = \lambda_0$ for every $o \in F$. For $N \in \mathscr{N}(S,T)$, let o^N be the solution defined by $o^N_i = 1$ for $i \in (V_{BR} \backslash (S \cup T)) \cup N$, and $o^N_i = 0$ otherwise. Since N is an $\mathscr{N}$-set, then o^N is feasible. Besides, if N is perfect, then $v_i = |N|$ for all $i \in N$, and so $\sum_{i \in S \cup T} \frac{o^N_i}{v_i} = \sum_{i \in N} \frac{1}{|N|} = 1$. Therefore, $o^N \in F$ whenever N is perfect.

Now, observe that $G^k_{S,T}$ is connected, for every $k \in N$. If $k > 1$, (F1) holds by hypothesis. If $k = 1$, it is implied by the fact that $G^1_{S,T}$ is complete or empty.

Let $i, j \in S \cup T$ such that $v_i = v_j = k \in \mathbb{N}$. Then, $i, j \in V^k_{S,T}$. If $ij \in E^k_{S,T}$, then there are two perfect $\mathscr{N}$-sets N_i and N_j such that $i \in N_i$, $j \in N_j$ and $N_i \triangle N_j = \{i, j\}$. Then, $o^{N_i} \in F$ and $o^{N_j} \in F$. Besides, these two solutions only differ in the variables o_i and o_j, and so we have $\lambda_i = \lambda_j$. If $ij \notin E^k_{S,T}$, we still get $\lambda_i = \lambda_j$ since $G^k_{S,T}$ is connected.

Now, let $i, j \in S \cup T$ such that $v_i \neq v_j$, $v_i, v_j \in \mathbb{N}$. Let $k \in \mathbb{N}$ such that $i \in V^k_{S,T}$. If i has at least one neighbor in $G^k_{S,T}$, then the hypothesis (F1) ensures the existence of a perfect $\mathscr{N}$-set N_i with $i \in N_i$ and $N_i \subseteq V^k_{S,T}$. On the other hand, if i is an isolated vertex in $G^k_{S,T}$, then the hypothesis (F1) ensures $v_i = k = 1$, hence $N_i = \{i\}$ is a perfect $\mathscr{N}$-set. In any case, we have a perfect $\mathscr{N}$-set N_i such that $|N_i| = v_i$ and $\lambda_l = \lambda_i$ for all $l \in N_i$. The same argument ensures the existence of a perfect $\mathscr{N}$-set N_j such that $|N_j| = \nu_j$ and $\lambda_l = \lambda_j$ for all $l \in N_j$. This implies that $o^{N_i} \in F$ and $o^{N_j} \in F$, so $\lambda^T o^{N_i} = \lambda^T o^{N_j}$, hence $\nu_i\lambda_i = \nu_j\lambda_j$.

Let $i \in V_B \backslash S$. By hypothesis (F2), let $N \in \mathscr{N}^*(S,T)$ be such that $(S \cup \{i\} \backslash N, T \backslash N)$ is linearly separable. This last condition implies that $o^N - e^i$ is a feasible solution, where e^i is the i-th unit vector. Since N is perfect, we have $o^N \in F$, and so $o^N - e^i \in F$ because $i \notin S \cup T$. Hence, we get $\lambda_i = 0$. A similar argument allows us to conclude that $\lambda_j = 0$ for every $j \in V_R \backslash T$.

This implies that λ is a multiple of the coefficient vector of inequality (4.6), which then defines a facet of P_1. ■

Let us call $\mathscr{N}$-set *elementary inequality* an $\mathscr{N}$-set inequality (4.6) related to a linearly inseparable pair (S,T) where $\nu_i = 1$ for all $i \in S \cup T$. In this case, we also say that (S,T) is a linearly inseparable elementary pair.

For $\mathscr{N}$-set elementary inequalities, we can specialize Theorem 4.2.2.

Corollary 4.2.3 *Let $(S \subseteq V_B, T \subseteq V_R)$ be a linearly inseparable elementary pair. The $\mathscr{N}$-set elementary inequality induced by (S,T) is facet-defining for P_1 if, and only if,*

(F_E) for each $j \in V_B \backslash S$ (resp. $j \in V_R \backslash T$), there exists $l \in S \cup T$ such that $(S \cup \{j\} \backslash \{l\}, T \backslash \{l\})$ (resp. $(S \backslash \{l\}, T \cup \{j\} \backslash \{l\})$) is linearly separable.

Proof Consider a linearly inseparable elementary pair (S,T). First, assume that (S,T) satisfies (F_E). Since $\nu_i = 1$ for all $i \in S \cup T$, we have $V^k_{S,T} = \emptyset$ for all $k > 1$. Therefore, (F1) holds. In addition, (F_E) directly implies (F2). Thus, the $\mathscr{N}$-set elementary inequality is facet-defining.

Now, suppose that (S,T) does not satisfy (F_E). Without loss of generality, we can then assume that there is $i \in V_B \setminus S$ such that $(S' \setminus \{l\}, T \setminus \{l\})$, for all $l \in S \cup T$, is linearly inseparable, where $S' = S \cup \{i\}$. This means that (S',T) is linearly inseparable, and

$$\nu_j' := \min\{|N| \mid N \in \mathscr{N}(S',T), j \in N\} = 2, \qquad \forall j \in S' \cup T.$$

Indeed, $\nu_j' \geq 2$ because (S,T) and (S',T) are linearly inseparable. On the other hand, since $\nu_j = 1$ for all $j \in S \cup T$, $\{i,j\}$ is an $\mathscr{N}$-set containing j and i, which implies $\nu_i' \leq 2$ and

$v_j^{'} \leq 2$ for all $j \in S \cup T$. Therefore, $v_j^{'} = 2$ for all $j \in S \cup T \cup \{i\}$, and so the inequality

$$\sum_{j \in S \cup T \cup \{i\}} o_j \geq 2$$

is valid for P_1. This inequality together with $o_i \geq 1$ dominate the elementary inequality given by (S,T), which then does not define a facet of P_1. ∎

4.2.1 ILP1's constraints

In this subsection, we derive facet-defining conditions for the constraints of ILP1. We start with the bounding constraints.

The solutions used in the proof of Proposition 4.1.3 allow us to show that they induce facets of P_1.

Proposition 4.2.4 *For every $i \in V_{BR}$, $o_i \geq 0$ and $o_i \leq 1$ are facet-defining for P_1.*

Proof Let $i \in V_{BR}$. The face defined by $o_i \leq 1$ contains the affinely independent points e and $e - e^j$, for all $j \in V_{BR} \setminus \{i\}$. The face defined by $o_i \geq 0$ contains the affinely independent points $e - e^i$ and $e - e^i - e^j$, for all $j \in V_{BR} \setminus \{i\}$. ∎

We now relate constraints (4.4) and (4.5) with $\mathcal{N}$-set elementary inequalities.

Proposition 4.2.5 *Constraints* (4.4) *and* (4.5) *are exactly all $\mathcal{N}$-set elementary inequalities.*

Proof First, we prove that every constraint in (4.4) and (4.5) is an $\mathcal{N}$-set elementary inequality. Consider a constraint in (4.4) related to i and S. Since $i \in H[S]$, then $(S,\{i\})$ is linearly inseparable. Besides, since S is minimal with regard to i i, it implies that $i \notin H[S \backslash \{u\}]$ for every $u \in S$. Then, $H[S \setminus \{u\}] \cap \{i\} = \emptyset$, $S \setminus \{u\} \cap H[\{i\}] = \emptyset$, and $H[S \backslash \{u\}] \cap H[\{i\}] \cap V_N = \emptyset$ for every $u \in S$. This means that, $(S \backslash \{u\}, \{i\})$ is linearly separable. Also, $(S, \emptyset)$ is linearly separable. Therefore, $\{i\}$ and $\{u\}$ for every $u \in S$ are minimal $\mathcal{N}$-sets. Thus, $v_j = 1$ for all $j \in S \cup \{i\}$.

Now, consider (4.5) related to S and T. As $S \cup T$ is minimal with regard to $H[S] \cap H[T] \cap V_N \neq \emptyset$, we have that, for every $j \in S \cup T$, $H[S \backslash \{j\}] \cap H[T \backslash \{j\}] \cap V_N = \emptyset$. Since $H[S] \cap T = H[T] \cap S = \emptyset$. It implies that $(S \backslash \{j\}, T \backslash \{j\})$ is linearly separable for every $j \in S \cup T$. Therefore, $v_j = 1$ for all $j \in S \cup T$.

Now, let $(S \subseteq V_B, T \subseteq V_R)$ be linearly inseparable and $v_j = 1$ for every $j \in S \cup T$. The corresponding $\mathscr{N}$-set inequality is $\sum_{j \in S \cup T} o_j \geq 1$. Since (S,T) is linearly inseparable, then $H[S] \cap T \neq \emptyset$ or/and $H[T] \cap S \neq \emptyset$ or/and $H[S] \cap H[T] \cap V_N \neq \emptyset$. If $H[S] \cap T \neq \emptyset$, since $v_j = 1$ for all $j \in T$, then $|T| \leq 1$; otherwise there would be $u \in T$ with $v_u > 1$. Also, since $v_j = 1$ for all $j \in S$, $(S \backslash \{j\}, T)$ is linearly separable and then S is minimal with regard to $i \in H[S]$, where $T = \{i\}$. Similarly, if $H[T] \cap S \neq \emptyset$, then T is minimal with regard to $i \in H[T]$, $S = \{i\}$. In both cases, the $\mathscr{N}$-set inequality is one of the inequalities in (4.4). Finally, suppose that $H[S] \cap H[T] \cap V_N \neq \emptyset$, $H[S] \cap T = \emptyset$ and $H[T] \cap S = \emptyset$. Since $v_j = 1$ for all $j \in S \cup T$, $(S \backslash \{j\}, T \backslash \{j\})$ is linearly separable for every $j \in S \cup T$. In particular, $H[S \backslash \{j\}] \cap H[T \backslash \{j\}] \cap V_N = \emptyset$ for all $j \in S \cup T$, and so $S \cup T$ is minimal. Therefore, the $\mathscr{N}$-set inequality is one of the inequalities in (4.5). ■

We specialize the definition of $\mathscr{N}$-set elementary inequality for constraints (4.4) and (4.5). The first ones will be called V_{BR}-disjoint $\mathscr{N}$-set elementary inequalities, whereas the second ones will be called V_N-disjoint $\mathscr{N}$-set elementary inequalities.

From Proposition 4.2.5, we can conclude that ILP1 is a set covering formulation where the constraint matrix A has rows indexed by $M = \{(S,T) \mid S \subseteq V_B, T \subseteq V_R$ is a linearly inseparable elementary pair$\}$, columns indexed by V_{BR}, and $a_{(S,T),i} = 1$ if, and only if, $i \in S \cup T$. Using the definition of the set covering polytope from Subsection 2.3, we have that $P_1 = P_I(A)$. Thus, we can relate conditions (F'1) and (F'2) to (F_E) of Corollary 4.2.3.

Corollary 4.2.6 *The V_{BR}-disjoint and V_N-disjoint $\mathscr{N}$-set elementary inequalities,* (4.4) *and* (4.5) *respectively, are facet-defining if, and only if, (F_E) holds. Moreover, they are the only facet-defining inequalities for P_1 with integer coefficients and right-hand side equal to* 1.

Some special cases of constraints (4.4)-(4.5) are:

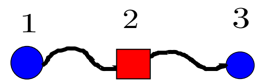
$$o_h + o_j + o_i \geq 1, \qquad \forall i \in V_{BR}, \forall \{h,j\} \subseteq V_{\bar{K}(i)} : i \in D_{hj}, \tag{4.7}$$

$$o_v + o_{v'} + o_w + o_{w'} \geq 1, \qquad \forall \{v,v'\} \subseteq V_B, \forall \{w,w'\} \subseteq V_R : D_{vv'} \cap D_{ww'} \cap V_N \neq \emptyset, \tag{4.8}$$

$$H[\{v,v'\}] \cap \{w,w'\} = H[\{w,w'\}] \cap \{v,v'\} = \emptyset.$$

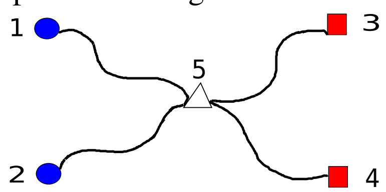
Constraints (4.7) are given by (4.4) where $|S| = 2$ and $i \in H[S]$ is replaced by the stronger condition $i \in D[S]$. They will be called *generalized* 3*-path constraints*. Inequalities (4.8) are constraints (4.5) with $|S| = |T| = 2$. They will be called *X-swing constraints*. They are easily separated in polynomial time. Besides, they are viable in practice whenever $H[\{l, p\}]$ is replaced by $D_{lp} \cup \{l, p\}$, for every $l, p \in V(G)$, in (4.8). Figures 13 and 14 show examples of them.

Figure 13 – An example of a generalized 3-path constraint.



$$o_1 + o_2 + o_3 \geq 1$$

Figure 14 – An example of X-swing constraint. Vertex 5 belongs to V_N.



$$o_1 + o_2 + o_3 + o_4 \geq 1$$

Applying Corollary 4.2.6 to (4.7) and (4.8) yields:

Corollary 4.2.7 *A constraint of type* (4.7) *defines a facet of* P_1 *if and only if* $u \notin H[\{h, j\}]$ *for every* $u \in V_{K(i)} \backslash \{i\}$*, or* $t \notin H[\{i, u\}]$ *for some* $t \in \{h, j\}$.

Corollary 4.2.8 *A constraint of type* (4.8) *defines a facet of* P_1 *if, and only if,*

1. *for every* $u \in V_B \backslash \{v, v'\}$*,* $(\{v, v', u\}, \{t\})$ *is linearly separable, for some* $t \in \{w, w'\}$*, or* $(\{t, u\}, \{w, w'\})$ *is linearly separable, for some* $t \in \{v, v'\}$*, and*
2. *for every* $u \in V_R \backslash \{w, w'\}$*),* $(\{t\}, \{w, w', u\})$ *is linearly separable, for some* $t \in \{v, v'\}$*, or* $(\{v, v'\}, \{t, u\})$ *is linearly separable, for some* $t \in \{w, w'\}$.

4.2.2 *Valid and facet-defining $\mathscr{N}$-set non-elementary inequalities for* P_1

In this subsection, we focus on $\mathscr{N}$-set non-elementary inequalities, i.e. $\mathscr{N}$-set inequalities different from the constraints of ILP1.

Proposition 4.2.9 *Let $S_B \subseteq V_B$ and $S_R \subseteq V_R$ such that* $\min\{|S_B|,|S_R|\} = l \geq 2$*. If $S_R \subseteq H[\{i,i'\}]$, for every $i,i' \in S_B$, $i \neq i'$, and $S_B \subseteq H[\{j,j'\}]$, for every $j,j' \in S_R$, $j \neq j'$, thus, the following inequality is valid for P_1:*

$$\sum_{i \in S_B \cup S_R} o_i \geq l. \tag{4.9}$$

Proof Let o be a feasible solution. Suppose by contradiction that o uses less than l outliers. By hypothesis, $\min\{|S_B|,|S_R|\} = l \geq 2$, which implies that there are at least 2 non-outliers in S_B or S_R. By symmetry, we can assume that $o_i = o_{i'} = 1$ for some $i,i' \in S_B$. Since $S_R \subseteq H[\{i,i'\}]$, constraints (4.1) imply that $o_j = 1$, for all $j \in S_R$. So, there are at least $|S_R| \geq l$ outliers: a contradiction. ■

Fortunately, only inequalities with $|S_B| = |S_R| = 2$ are non-redundant:

Proposition 4.2.10 *Inequalities (4.9) are dominated by inequalities with $|S_B| = |S_R| = 2$.*

Proof Without loss of generality, suppose that $|S_B| = l \geq 2$, $|S_R| = m \geq 2$, and $l \leq m$. Then, there are $\binom{l}{2} \cdot \binom{m}{2}$ inequalities involving only two vertices of each side. Summing up all these inequalities yields the following expression:

$$\begin{aligned}
& (l-1)\binom{m}{2}(\sum_{i \in S_B} o_i) + (m-1)\binom{l}{2}(\sum_{i \in S_R} o_i) \geq 2\binom{l}{2}\binom{m}{2} \\
\Rightarrow\ & (l-1)(\sum_{i \in S_B} o_i) + \frac{(m-1)\binom{l}{2}}{\binom{m}{2}}(\sum_{i \in S_R} o_i) \geq 2\binom{l}{2} \\
\Rightarrow\ & (l-1)(\sum_{i \in S_B} o_i) + \frac{l(l-1)}{m}(\sum_{i \in S_R} o_i) \geq 2\binom{l}{2} \\
\Rightarrow\ & \sum_{i \in S_B} o_i + \frac{l}{m}(\sum_{i \in S_R} o_i) \geq l,
\end{aligned}$$

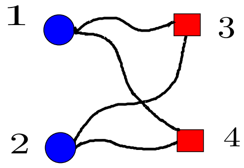
which is equal or stronger than inequality (4.9), as $l \leq m$. ■

So, we just look for inequalities where $|S_B| = |S_R| = 2$, which we call *generalized C_4 inequalities*. They become (see the example of Figure 15):

$$\sum_{i \in S_B \cup S_R} o_i \geq 2. \tag{4.10}$$

The validity of (4.10) can also be proved using Proposition 4.2.1. Note that $\nu_i = 2$ for every $i \in S_B \cup S_R$, and thus (4.10) is an $\mathscr{N}$-set inequality. Moreover, it always defines a facet of P_1, as showed by the proposition below.

Figure 15 – C_4 inequality example.



$o_1 + o_2 + o_3 + o_4 \geq 2$

Proposition 4.2.11 *Inequalities* (4.10) *define facets of* P_1.

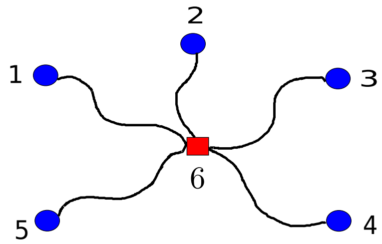
Proof Let $|S_B| = |S_R| = 2$ with $S_B = \{v, v'\}$ and $S_R = \{w, w'\}$ such that they define an inequality of type (4.10). Observe that any subset $\{i, j\} \subset \{v, v', w, w'\}$ defines a perfect $\mathscr{N}$-set. So, $ij \in E^2_{S_B,S_R}$ for every $i, j \in \{v, v', w, w'\}$. Since $\nu_i = 2$ for every $i \in S_B \cup S_R$, the safe graph $G^2_{S_B,S_R} = (S_B \cup S_R, E^2_{S_B,S_R})$ is complete. Thus, the first condition of Theorem 4.2.2 holds for (4.10).

Now, let $i \in V_B \backslash \{v, v'\}$. Then, $N = \{w, w'\}$ is a perfect $\mathscr{N}$-set such that $(\{v, v', i\} \backslash N, \{w, w'\} \backslash N)$ is linearly separable. Similarly, for $j \in V_R \backslash \{w, w'\}$, $N = \{v, v'\}$ is a perfect $\mathscr{N}$-set such that $(\{v, v'\} \backslash N, \{w, w', j\} \backslash N)$ is linearly separable. Therefore, the second condition of Theorem 4.2.2 holds for (4.10). ■

We anticipate that the generalized C_4 inequalities were quite useful to reduce the computational time of the ILP1 formulation (see Chapter 6 for details). It is worth noting that these inequalities result from a configuration of geodesic convex combinations that cannot occur in the Euclidean space.

An $\mathscr{N}$-set inequality induced by a generalized star is shown in Figure 16. These inequalities are formally defined in the following proposition and will be called *star tree* inequalities.

Figure 16 – Example of a star tree inequality for ILP1.



$$o_1 + o_2 + o_3 + o_4 + o_5 + 4o_6 \geq 4$$

Proposition 4.2.12 *Let $K \in \{B,R\}$, $L \subseteq V_K$ and $i \in V_{\bar{K}}$ be such that $i \in H[\{h,j\}]$ for all $h,j \in L$, $h \neq j$. The following star tree inequality is valid for P_1:*

$$\sum_{h \in L} o_h + (|L|-1)o_i \geq (|L|-1). \tag{4.11}$$

Moreover, if for every $u \in V_{\bar{K}} \backslash \{i\}$, $u \notin H[L]$ or $j \notin H[\{i,u\}]$, for some $j \in L$, then (4.11) *defines a facet of P_1.*

Proof The pair $(\{i\},L)$ is linearly inseparable. Besides, it is easy to see that $v_i = 1$ and $v_h = |L|-1$ for all $h \in L$. By Proposition 4.2.1, $\sum_{h \in L} o_h + (|L|-1)o_i \geq (|L|-1)$ is valid for P_1. Also, note that $G^1_{\{i\},L} = (\{i\},\emptyset)$ and $G^{|L|-1}_{\{i\},L} = (L, E^{|L|-1})$ are complete safe graphs, since any $N \subset L$ with $|N| = |L|-1$ is a perfect $\mathscr{N}$-set. Thus, condition 1 of Theorem 4.2.2 holds for (4.11).

Since $v_i = 1$, $N = \{i\}$ is a perfect $\mathscr{N}$-set such that $(\{i\}\backslash N, L \cup \{w\}\backslash N)$ is linearly separable for every $w \in V_K \backslash L$. By hypothesis, for every $u \in V_{\bar{K}} \backslash \{i\}$, $u \notin H[L]$ or $j \notin H[\{i,u\}]$, for some $j \in L$. So, there is a perfect $\mathscr{N}$-set $N \subset (L \cup \{i\})$ such that $(\{i,u\}\backslash N, L\backslash N)$ is linearly separable for every $u \in V_{\bar{K}} \setminus \{i\}$. Therefore, condition (F2) of Theorem 4.2.2 holds. ■

Next, we show another $\mathscr{N}$-set inequality to be called *generalized alternating path inequality* or *alternating path inequality* for short. These inequalities are inspired by a family of facet-defining inequalities defined by (LIMA, 2011) for the convex recoloring problem.

Proposition 4.2.13 *Let $S_{hj} =< h = l_1, q_1, \ldots, l_t, q_t, j >$, $t \leq \lfloor \delta(h, j) \rfloor$, be a sequence of distinct vertices of V_{BR} with odd cardinality that corresponds to an incomplete or complete shortest path from h to j in G. Let the vertices along the sequence have alternating classes. So, the alternating path inequality is valid for P_1:*

$$\sum_{i \in V(S_{hj})} o_i \geq t. \tag{4.12}$$

Proof This proof is adapted from (LIMA, 2011). We use induction on t. Note that each triple $\phi = (h, j, t)$ can induce one or more inequalities. See Figure 17. When $t = 1$, the corresponding inequalities are exactly the ones in (4.7), which are already in the formulation, and so are valid. Now, let $t = 2$. Consider the sum of the following three valid inequalities for $t = 1$:

$$\begin{array}{cccccccc}
o_{l_1} & +o_{q_1} & +o_{l_2} & & & \geq & 1 \\
 & & +o_{l_2} & +o_{q_2} & +o_j & \geq & 1 \\
+o_{l_1} & & & +o_{q_2} & +o_j & \geq & 1 \\
\hline
2o_{l_1} & +o_{q_1} & +2o_{l_2} & +2o_{q_2} & +2o_j & \geq & 3
\end{array}$$

Thus, we have that inequality $2o_{l_1} + o_{q_1} + 2o_{l_2} + 2o_{q_2} + 2o_j \geq 3$ is valid for P_1. Summing this inequality with the inequality $o_{q_1} \geq 0$, we get:

$$2o_{l_1} + 2o_{q_1} + 2o_{l_2} + 2o_{q_2} + 2o_j \geq 3.$$

Then, dividing the inequality above by the common coefficient 2 yields:

$$o_{l_1} + o_{q_1} + o_{l_2} + o_{q_2} + o_j \geq \frac{3}{2}.$$

Since we want integer valued solutions for the variables o, we can round up the right-hand side to get the following stronger valid inequality for P_1:

$$o_{l_1} + o_{q_1} + o_{l_2} + o_{q_2} + o_j \geq 2.$$

Therefore, all generalized alternating path inequalities for $\phi = (h, j, 2)$ are valid for P_1. They are obtained by the Chvátal-Gomory procedure.

For a better illustration of the combination and the round up procedure, let us show the validity proof for the generalized alternating path inequalities for $\phi = (h, j, 3)$. Consider the sum of the following four valid inequalities:

$$
\begin{array}{cccccccccc}
+o_{l_1} & +o_{q_1} & +o_{l_2} & +o_{q_2} & +o_{l_3} & & & \geq & 2 \\
 & & +o_{l_2} & +o_{q_2} & +o_{l_3} & +o_{q_3} & +o_j & \geq & 2 \\
+o_{l_1} & & & & & +o_{q_3} & +o_j & \geq & 1 \\
 & +o_{q_1} & & & & & & \geq & 0 \\
\hline
2o_{l_1} & +2o_{q_1} & +2o_{l_2} & +2o_{q_2} & +2o_{l_3} & +2o_{q_3} & +2o_j & \geq & 5.
\end{array}
$$

Dividing the inequality above and rounding up the right-hand side yields:

$o_{l_1} + o_{q_1} + o_{l_2} + o_{q_2} + o_{l_3} + o_{q_3} + o_j \geq 3.$

Therefore, all generalized alternating path inequalities related to $\phi = (h, j, 3)$ are valid for P_1.

We can apply an analogous procedure to obtain the inequalities induced by $\phi(h, j, t)$, $t \geq 4$. Precisely, we sum up the following four inequalities:

- the inequality induced by $\phi(h = l_1, l_t, t-1)$;
- the inequality induced by $\phi(l_2, j, t-1)$;
- the inequality $o_{l_1} + o_{q_t} + o_j \geq 1$;
- the inequality $o_{q_1} \geq 0$.

After that, we divide by 2 the resulting inequality and round up the right-hand side. ■

Actually, the generalized alternating path inequalities are $\mathscr{N}$-set inequalities, as the following proposition states.

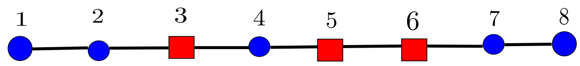
Proposition 4.2.14 *Inequalities* (4.12) *are $\mathscr{N}$-set inequalities.*

Proof Let $S_{hj} = < h = l_1, q_1, \ldots, l_t, q_t, l_{t+1} = j >$ as in Proposition 4.2.13. Then, $(\{l_1, \ldots, l_{t+1}\}, \{q_1, \ldots, q_t\})$ is linearly inseparable. Besides, $\nu_u \geq t$, for every $u \in V(S_{hj})$, since (4.6) is a valid inequality. Observe that $N = \{q_j \mid j \in \{1, \ldots, t\}\}$ is an $\mathscr{N}$-set such that N contains q_i and $|N| = t$, for every $i \in \{1, \ldots, t\}$. Since $\nu_{q_i} \geq t$, we have $\nu_{q_i} = t$, for every $i \in \{1, \ldots, t\}$.

We can note that $N = (\{l_u \mid u \in \{1, \ldots, t\}\} \cup \{j\}) \backslash \{w\}$ is an $\mathscr{N}$-set if $w = h = l_1$ or $w = j$ (the extreme points of S_{hj}). Moreover, $|N| = t$. So, for every $i \in \{1, \ldots, t\}$, there is an $\mathscr{N}$-set N such that N contains l_i and $|N| = t$, which implies that $\nu_{l_i} = t$. The same applies for j, resulting in $\nu_j = t$.

Therefore, $\nu_u = t$, for every $u \in V(S_{hj})$ leads to the $\mathscr{N}$-set inequality (4.12). ■

Figure 17 – An example of alternating path inequality for ILP1.



$$o_1 + o_3 + o_4 + o_6 + o_8 \geq 2$$

4.2.3 A compact formulation for the 2-SGC problem

The second integer linear formulation is obtained by adding some variables to ILP1 so as to reduce the number of constraints to a polynomial order. The new variables, z, are used to determine if a vertex belongs to the convex hull of the non-outliers of a given class. So, we have for each $i \in V_{BR}$,

$$o_i = \begin{cases} 1, & \text{if } i \in V_{BR} \setminus (A_B \cup A_R), \\ 0, & \text{if } i \in (A_B \cup A_R), \end{cases}$$

and for each $K \in \{B,R\}$ and $i \in V$,

$$z_{Ki} = \begin{cases} 1, & \text{if } i \in H[A_K], \\ 0, & \text{otherwise.} \end{cases}$$

The formulation is defined by

$$\text{(ILP2)} \ \min \sum_{i \in V_{BR}} o_i \tag{4.13}$$

$$\text{st: } o_i \geq z_{\bar{K}(i)i}, \qquad \forall i \in V_{BR}, \tag{4.14}$$

$$z_{Bi} + z_{Ri} \leq 1, \qquad \forall i \in V_N, \tag{4.15}$$

$$z_{K(i)i} + o_i \geq 1, \qquad \forall i \in V_{BR}, \tag{4.16}$$

$$z_{Kh} + z_{Kj} - z_{Ki} \leq 1, \qquad \forall K \in \{B,R\}, \forall h,i,j \in V : i \in D_{hj}, \tag{4.17}$$

$$o \in \mathbb{B}^{|V_{BR}|}, z \in \mathbb{B}^{2|V|}. \tag{4.18}$$

Formulations ILP1 and ILP2 can be related as follows.

Proposition 4.2.15 *Let $F1$ and $F2$ be the feasible sets of ILP1 and ILP2, respectively. Then,* $F1 = \text{proj}_o(F2)$.

Proof Let $o \in \mathbb{B}^{|V_{BR}|}$ be a feasible solution of ILP1. As in the proof of Proposition 4.1.1, let $A_K = \{j \in V_K : o_j = 0\}$, for $K \in \{B,R\}$. By that proposition, A_B and A_R satisfy conditions (C1)-(C3). Define $z \in \mathbb{B}^{2|V|}$ such that $z_{Ki} = 1$ if, and only if, $i \in H[A_K]$, for all $K \in \{B,R\}$ and $i \in V$. We have to show that (o,z) is feasible for ILP2. To check constraints (4.14) and (4.16), let $i \in V_{BR}$

and suppose that $o_i = 0$ (otherwise they are trivially satisfied). Then, $i \in A_{K(i)} \subseteq H[A_{K(i)}]$, and so $z_{K(i)i} = 1$. Besides, (C1) and (C2) imply that $i \notin H[A_{\bar{K}(i)}]$, and so $z_{\bar{K}(i)i} = 0$. These show that constraint (4.16) and (4.14) are satisfied by (o,z). Now let $i \in V_N$. By (C3), $i \notin H[A_B] \cap H[A_R]$. In other terms, (4.15) is satisfied. Finally, by the geodesic convexity definition, for each $K \in \{B,R\}$ and $i \in D_{hj}$, if $z_{Kh} = z_{Kj} = 1$ then we must have $z_{Ki} = 1$. This shows that constraints (4.17) are satisfied. Therefore, $F1 \subseteq \text{proj}_o(F2)$.

Conversely, let $(o,z) \in F2$. We have to show that o satisfies (4.1)-(4.2). First, let $i \in V_{K(i)}$ and $S \subseteq V_{\bar{K}(i)}$ such that $i \in H[S]$. Suppose that $o_j = 0$ for all $j \in S$; otherwise (4.1) is trivially satisfied. By (4.16), $z_{\bar{K}(i)j} = z_{\bar{K}(j)j} = 1$ for all $j \in S$. Then, since $i \in H[S]$, we can use (4.17) to conclude that $z_{\bar{K}(i)i} = 1$. So, (4.14) ensures that $o_i = 1$, showing that (4.1) is satisfied. Now, let $S \subseteq V_B$ and $T \subseteq V_R$ such that there is $i \in H[S] \cap H[T] \cap V_N$. Suppose by contradiction that (4.2) is violated, i.e. $o_j = 0$ for all $j \in S \cup T$. By (4.16), it follows that $z_{Rj} = 1$ for all $j \in S$, and $z_{Bj} = 1$ for all $j \in T$. Since $i \in H[S] \cap H[T]$, (4.17) implies that $z_{Ri} = z_{Bi} = 1$, which contradicts (4.15). Therefore, o satisfies (4.2). Then, we conclude that $o \in F2$. ■

Propositions 4.1.1 and 4.2.15 imply the correctness of ILP2.

Corollary 4.2.16 *Formulation ILP2 is correct.*

Although the feasible sets of ILP1 and ILP2 are related as in Proposition 4.2.15, the same does not occur with their linear relaxations. Actually, we can find examples where the linear relaxation of ILP1 provides a better bound than ILP2, and vice-versa.

4.2.4 ILP2's polyhedron

Let $P_2 = conv(\{(o,z) \in \mathbb{B}^{|V_{BR}| \times \mathbb{B}^{2|V|}} \mid (o,z) \text{ satisfies (4.14)-(4.17)}\})$. Observe that there are $n_o = |V_{BR}|$ of variables o and $n_z = |C_{BR}||V|$ of variables z, totalizing $n_2 = |V_{BR}| + (|C_{BR}||V|)$ variables.

Proposition 4.2.17 P_2 *is a full-dimensional polyhedron.*

Proof Consider the following n_2+1 points in $\mathbb{R}^{n_2}$:

$$v^{Ki} = (e, e^{ki}) = \begin{array}{|ccc|c|} o & z & z_{Ki} & z \\ \hline 1\dots1 & 0\dots0 & 1 & 0\dots0 \\ \hline \end{array}, \quad i \in V,\, K \in \{B,R\},$$

$$w^0 = (e,0) = \begin{array}{|c|c|} o & z \\ \hline 1\dots1 & 0\dots0 \\ \hline \end{array},$$

$$w^i = (e-e^i, e^{K(i)i}) = \begin{array}{|ccc|ccc|} o & o_i & o & z & z_{K(i)i} & z \\ \hline 1\dots1 & 0 & 1\dots1 & 0\dots0 & 1 & 0\dots0 \\ \hline \end{array}, \quad i \in V_{BR}.$$

They clearly are feasible solutions of ILP2. To show that they are affinely independent points, let $\alpha_{v^{Ki}}$, and β_{w^i} be coefficients such that $(\sum_{K\in\{B,R\}}\sum_{i\in V}\alpha_{v^{Ki}}v^{Ki}) + \sum_{i\in V_{BR}}\beta_{w^i}w^i + \sum_{i\in V_{BR}}\beta_{w^0}w^0 = 0$ and $(\sum_{K\in\{B,R\}}\sum_{i\in V}\alpha_{v^{Ki}}) + \sum_{i\in V_{BR}}\beta_{w^i} + \beta_{w^0} = 0$. Observe that $o_i = 0$ occurs only in point w^i, so, since $(\sum_{K\in\{B,R\}}\sum_{i\in V}\alpha_{v^{Ki}}) + \sum_{i\in V_{BR}}\beta_{w^i} + \beta_{w^0} = 0$, we can conclude that $\beta_{w^i} = 0$, for all $i \in V_{BR}$. Thus, ignoring these points, we can note that $z_{Ki} = 1$ only occurs in point v^{Ki}, for $i \in V$, $K \in \{B,R\}$. So, $\alpha_{v^{Ki}} = 0$, for all $i \in V$ and $K \in \{B,R\}$, which implies that $\beta_{w^0} = 0$. Then, all coefficients are equal to 0.

Therefore, P_2 is a full-dimensional polyhedron. ■

By Propositions 2.2.14 and 4.2.15, we have that:

Proposition 4.2.18 $P_1 = \mathrm{proj}_o(P_2)$.

Propositions 2.2.10 and 4.2.18 ensure that all valid inequalities derived for ILP1 in the previous sections are also valid for ILP2. In general, it follows that:

Corollary 4.2.19 *If $\pi^T o \geq \pi_0$ is valid P_1, then it is valid for P_2.*

However, the facetness conditions for P_1 are not directly transferred to P_2, even for the bounding inequalities. Anyway, just as in P_1, the o-coefficients of all facet-defining inequalities for P_2, except for the bounding inequalities, are non-negative.

Proposition 4.2.20 *Let $\pi^T o + \mu^T z \geq \pi_0$ be a facet-defining inequality of P_2. If it is different from $o_i \leq 1$, for all $i \in V_{BR}$, then $\pi \geq 0$ and $\pi_0 > 0$.*

Proof Let $i \in V_{BR}$. Since the inequality is different from $o_i \leq 1$, the facet $F := \{(o,z) \in P_2 : \pi^T o + \mu^T z = \pi_0\}$ contains a point $(\bar{o},\bar{z})$ with $\bar{o}_i = 0$. In addition, the point $(o',\bar{z})$ such that $o'_j = \bar{o}_j$, for all $j \neq i$, and $o'_i = 1$ belongs to P_2. Therefore, $\pi^T(o'-\bar{o}) + \mu^T(z-z) \geq \pi_0 - \pi_0 = 0$, which leads to $\pi_i \geq 0$. So, $\pi \geq 0$. Since $o_i \geq z_{K(i)i} \geq 0$ are valid, $o_i \geq 0$ can not be facet-defining. Then, there is $(\hat{o},z) \in F$ with $\hat{o}_i = 1$. Since $\pi \neq 0$ and i is an arbitrary vertex in V_{BR}, we can assume that $\pi_i > 0$. Thus, we get $\pi_0 = \pi^T \hat{o} \geq \pi_i > 0$. ■

4.2.5 ILP2's constraints

In this subsection, we focus on facet-defining conditions for the constraints of ILP2.

The solutions used in the proof of Proposition 4.2.17 also allow us to show the following facetness results for the bounding inequalities:

Proposition 4.2.21 *For every $i \in V_{BR}$, $o_i \leq 1$ is facet-defining for P_2.*

Proof The points $(e,0)$, $(e-e^j, e^{K(j)j})$ for every $j \in V_{BR} \setminus \{i\}$, and (e, e^{Kj}) for every $j \in V$ and $K \in K \in \{B,R\}$, used in the proof of Proposition 4.2.17, all lie in the face defined by $o_i \leq 1$, and so it is actually a facet of P_2. ■

Proposition 4.2.22 *Let $i \in V$ and $K \in \{B,R\}$. If $i \in V_N$, or $i \in V_{BR}$ and $K = \bar{K}(i)$, then inequality $z_{Ki} \geq 0$ is facet-defining for P_2.*

Proof Let $i \in V$, $K \in \{B,R\}$, and $F = \{(o,z) \in P_2 : z_{Ki} = 0\}$. If $i \in V_N$, or $i \in V_{BR}$ and $K = \bar{K}(i)$, the points $(e,0)$, $(e-e^j, e^{K(j)j})$ for every $j \in V_{BR}$, and (e, e^{Kj}) for every $j \in V \setminus \{i\}$ and $K \in \{B,R\}$ belong to F. Then, F is a facet. ■

Proposition 4.2.23 *For every $i \in V_{BR}$, $z_{K(i)i} \leq 1$ is facet-defining for P_2.*

Proof Suppose that $i \in V_{BR}$ and $K = K(i)$. Let $F = \{(o,z) \in P_2 : z_{Ki} = 1\}$ and assume that $F \subseteq F' := \{(o,z) \in P_2 : \pi^T o + \mu^T z = \pi_0\}$. It is enough to prove that $\pi = 0$, $\mu_{\overline{K}j} = 0$ for all $j \in V$, and $\mu_{Kj} = 0$ for all $j \in V \setminus \{i\}$. We consider the following cases (in each of them we present two points in $F \subseteq F'$ to get the desired result):

- $\mu_{\overline{K}j} = 0$ for all $j \in V$: The points $(e, e^{Ki}) \in F$ and $(e, e^{Ki} + e^{\overline{K}j}) \in F$ imply that $\mu^T e^{\overline{K}j} = \mu_{\overline{K}j} = 0$. Note that the second point belongs to F even if $j = i$;
- $\mu_{Kj} = 0$ for all $j \in V \setminus \{i\}$: We apply induction on d_j, the distance from i to j in G. For $d \geq 1$, let $V_d = \{j \in V \setminus \{i\} : d_j \leq d\}$. If $j \in V_1$, we use the points $(e, e^{Ki}) \in F$ and $(e, e^{Ki} + e^{Kj}) \in F$ to get $\mu_{Kj} = 0$. Suppose that $\mu_{Kj} = 0$ for all $j \in V_d$, for some $d \geq 1$. Now, consider $j \in V_{d+1}$. The point $(e, \sum_{\ell \in H[\{i,j\}]} e^{K\ell})$ belongs to F. Since $H[\{i,j\}] \setminus \{i,j\} \subseteq V_d$, we have that $\mu_{K\ell} = 0$, for all $\ell \in H[\{i,j\}] \setminus \{i,j\}$. Using such a point and $(e, e^{Ki}) \in F$, we conclude that $\mu^T e^{Kj} = \mu_{Kj} = 0$.
- $\pi_j = 0$ for all $j \in V_{\overline{K}}$: The points $(e, e^{Ki}) \in F$ and $(e - e^j, e^{Ki} + e^{\overline{K}j}) \in F$ lead to $\pi_j = 0$.
- $\pi_j = 0$ for all $j \in V_K$: The points $(e, e^{Ki}) \in F$ and $(e - e^j, \sum_{\ell \in H[\{i,j\}]} e^{K\ell}) \in F$ show that $\pi_j = \sum_{\ell \in H[\{i,j\}] \setminus \{i\}} \mu_{K\ell} = 0$. ■

Proposition 4.2.24 *For every $i \in V_N$, $z_{Bi} + z_{Ri} \leq 1$ is facet-defining for P_2.*

Proof Consider the following n_2 points:

$$v^{Ki} = (e, e^{Ki}) = \begin{array}{|ccc|c|} o & z_{Ki} & z_{\bar{K}i} & z \\ \hline 1\ldots1 & 1 & 0 & 0\ldots0 \\ \hline \end{array}, \quad K \in \{B,R\},$$

$$v^{Kj} = (e, e^{\bar{K}i + e^{Kj}}) = \begin{array}{|cccc|c|} o & z_{Ki} & z_{\bar{K}i} & z_{Kj} & z \\ \hline 1\ldots1 & 0 & 1 & 1 & 0\ldots0 \\ \hline \end{array}, \quad j \in V \backslash \{i\},\ K \in \{B,R\},$$

$$v^{j} = (e - e^j, e^{\bar{K}i + e^{Kj}}) = \begin{array}{|cccc|cc|} o_j & o & z_{Ki} & z_{\bar{K}i} & z_{Kj} & z \\ \hline 0 & 1\ldots1 & 0 & 1 & 1 & 0\ldots0 \\ \hline \end{array}, \quad K \in \{B,R\},\ j \in V_K \backslash \{i\}.$$

They clearly are feasible solutions of ILP2 and satisfy $z_{Bi} + z_{Ri} \leq 1$ at equality. To show that they are affinely independent points, let $(\sum_{K \in \{B,R\}} \sum_{j \in V \backslash \{i\}} \alpha_{v^{Kj}} v^{Kj}) + (\sum_{K \in \{B,R\}} \sum_{j \in V_K} \alpha_{v^j} v^j)$ $+ \alpha_{v^{Bi}} v^{Bi} + \alpha_{v^{Ri}} v^{Ri} = 0$ and $(\sum_{K \in \{B,R\}} \sum_{j \in V \backslash \{i\}} \alpha_{v^{Kj}}) + (\sum_{K \in \{B,R\}} \sum_{j \in V_K} \alpha_{v^j}) + \alpha_{v^{Bi}} + \alpha_{v^{Ri}} = 0$. Observe that $o_j = 0$ occurs only in point v^j, so $\alpha_{v_j} = 0, \forall j \in V_{BR}$. By ignoring such v_j points, we note that $z_{Kj} = 1$ occurs only in point v^{Kj}, for every $K \in \{B,R\}$ and $j \in V \backslash \{i\}$, hence $\alpha_{v^{Kj}} = 0$. Thus, we end up with $\alpha_{v^{Bi}} v^{Bi} + \alpha_{v^{Ri}} v^{Ri} = 0$, which yields $\alpha_{v^{Bi}} = \alpha_{v^{Ri}} = 0$. Therefore, they are all affinely independent points. ■

Proposition 4.2.25 *For every $i \in V_{BR}$, $o_i \geq z_{\bar{K}(i)i}$ is facet-defining for P_2.*

Proof Suppose, without loss of generality, that $i \in V_B$. Consider the following n_2 points:

$v^i = (e - e^i, e^{K(i)i}) =$

o_i	o	$z_{K(i)i}$	z
0	$1\ldots1$	1	$0\ldots0$

,

$v^{\bar{K}i} = (e, e^{\bar{K}(i)i}) =$

o	$z_{\bar{K}(i)i}$	z
$1\ldots1$	1	$0\ldots0$

,

$v^{Bj} = (e, e^{Ri} + e^{Bj}) =$

o	z_{Ri}	z_{Bj}	z
$1\ldots1$	1	1	$0\ldots0$

, $j \in V$.

$v^{Rj} = (e - e^i, e^{Bi} + e^{Rj}) =$

o_i	o	z_{Bi}	z_{Rj}	z
0	$1\ldots1$	1	1	$0\ldots0$

, $j \in V \backslash \{i\}$.

$v^{oBj} = (e - e^j, e^{Ri} + e^{Bj}) =$

o_j	o	z_{Ri}	z_{Bj}	z
0	$1\ldots1$	1	1	$0\ldots0$

, $j \in V_B \backslash \{i\}$.

$v^{oRj} = (e - e^j - e^i, e^{Bi} + e^{Rj}) =$

o_i	o_j	o	z_{Bi}	z_{Rj}	z
0	0	$1\ldots1$	1	1	$0\ldots0$

, $j \in V_R$.

Similarly to the proof of Proposition 4.2.24, we can conclude that these are n_2 affinely independent points satisfying $o_i \geq z_{\bar{K}(i)i}$ at equality. ■

Proposition 4.2.26 *For every $i \in V_{BR}$, $z_{K(i)i} + o_i \geq 1$ is facet-defining for P_2.*

Proof Suppose, without loss of generality, that $i \in V_B$. The n_2 points below are affinely independent and satisfy $z_{K(i)i} + o_i \geq 1$ at equality:

$$v^0 = (e,0) = \begin{array}{|c|c|} o & z \\ \hline 1\dots1 & 0\dots0 \\ \hline \end{array},$$

$$v^{Kj} = (e,e^{Kj}) = \begin{array}{|cc|c|} o & z_{Kj} & z \\ \hline 1\dots1 & 1 & 0\dots0 \\ \hline \end{array}, \quad j \in V,\ K \in \{B,R\},\ Kj \neq Bi,$$

$$v'^{Kj} = (e-e^j,e^{Kj}) = \begin{array}{|cc|cc|} o_j & o & z_{Kj} & z \\ \hline 0 & 1\dots1 & 1 & 0\dots0 \\ \hline \end{array}, \quad K \in \{B,R\},\ j \in V_K. \blacksquare$$

We close this subsection by showing the constraints of ILP2 that do not define facets of P_2. They complement the results of Propositions 4.2.21 and 4.2.23.

Proposition 4.2.27 *The constraints below do not define facets of* P_2*.*

1. $o_i \geq 0$*, for every* $i \in V_{BR}$*.*
2. $z_{\bar{K}(i)i} \leq 1$*, for every* $i \in V_{BR}$*.*
3. $z_{Ki} \leq 1$*, for every* $i \in V_N$ *and* $K \in \{B,R\}$*.*
4. $z_{K(i)i} \geq 0$*, for every* $i \in V_{BR}$*.*

Proof By (4.14), $o_i = 0$ implies $z_{\bar{K}(i)i} = 0$ and $z_{\bar{K}(i)i} = 1$ implies $o_i = 1$. So, $o_i \geq 0$ and $z_{\bar{K}(i)i} \leq 1$ can not define facet of P_2. Clearly, $z_{Ki} \leq 1$ is dominated by constraint $z_{Ki} + z_{\bar{K}i} \leq 1$. By (4.16), $z_{K(i)i} = 0$ implies $o_i = 1$. So, $z_{K(i)i} \geq 0$ can not define a facet of P_2. $\blacksquare$

4.2.6 Valid and facet-defining inequalities for P_2

In this subsection, we derive valid inequalities for P_2. Some of them are counterpart of inequalities presented for P_1. First, let us consider the star tree inequalities (4.11). Of course, they are valid for P_2 (see Corollary 4.2.19). However, we can replace variables o by variables z to set stronger inequalities, as shown below.

Proposition 4.2.28 *Let $i \in V$ and $L \subseteq V \setminus \{i\}$ be such that $i \in H[\{h,j\}]$ for all $h, j \in L$, $h \neq j$. The following inequalities are valid for P_2:*

$$\sum_{h \in L} z_{Kh} - (|L|-1)z_{Ki} \leq 1, \qquad K \in \{B,R\}. \tag{4.19}$$

Moreover, for $K = K(i)$ and $L \subseteq V_K$, (4.19) *dominates* (4.11).

Proof Let $K \in \{B,R\}$ and suppose that (o,z) is a feasible solution of ILP2. If $\sum_{h \in L} z_{Kh} \leq 1$, then inequality (4.19) trivially holds. So, suppose that $\sum_{h \in L} z_{Kh} > 1$. In this case, it means that there are at least two vertices in L, say j, j', such that $z_{Kj} = z_{Kj'} = 1$. By hypothesis, $i \in H[\{j, j'\}]$, and by constraint (4.17), $z_{Ki} = 1$. Therefore, $(|L|-1)z_{Ki} = (|L|-1) \geq \sum_{h \in L} z_{Kh} - 1$, and so $\sum_{h \in L} z_{Kh} - (|L|-1)z_{Ki} \leq 1$ holds.

Now, assume that $K = \bar{K}(i)$ and $L \subseteq V_K$, i.e. $K(j) = K$ for all $j \in L$. Then, (4.19) becomes

$$1 \geq \sum_{h \in L} z_{K(h)h} - (|L|-1)z_{\bar{K}(i)i}.$$

Using (4.14) and (4.16), we get

$$1 \geq \sum_{h \in L} (1 - o_h) - (|L|-1)o_i,$$

or still

$$\sum_{h \in L} o_h + (|L|-1)o_i \geq |L| - 1,$$

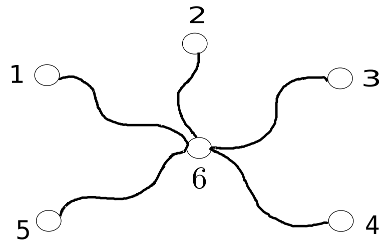
which is (4.11). ■

Figure 18 illustrates an example of a star tree inequality for vertex $i = 6$ and $L = \{1, \ldots, 5\}$. If the star tree is actually an induced star, then inequality (4.19) is a facet-defining for P_2.

Proposition 4.2.29 *Let $i \in V$ and $L \subseteq V \setminus \{i\}$ be such that $|L| \geq 2$, $(i,j) \in E(G)$, for all $j \in L$, and $(u,w) \notin E(G)$, for all $u, w \in L$. The following inequalities are facet-defining for P_2:*

$$\sum_{h \in L} z_{Kh} - (|L|-1)z_{Ki} \leq 1, \qquad K \in \{B,R\}. \tag{4.20}$$

Figure 18 – An example of star tree inequality for ILP2.



$$z_{Kv1} + z_{Kv2} + z_{Kv3} + z_{Kv4} + z_{Kv5} - 4z_{Kv6} \leq 1, \quad K \in \{B, R\}$$

Proof Let $K \in \{B,R\}$ and $F = \{(o,z) \in P_2 : \sum_{h \in L} z_{Kh} - (|L|-1)z_{Ki} = 1\}$. Assume that $F \subseteq F' := \{(o,z) \in P_2 : \pi^T o + \mu^T z = \pi_0\}$. We prove that $\pi^T o + \mu^T z = \pi_0$ is a multiple of $\sum_{h \in L} z_{Kh} - (|L|-1)z_{Ki} = 1$ by cases, as follows:

- $\mu_{\overline{K}j} = 0$ for all $j \in V$: The points $(e, e^{Kh}) \in F$ and $(e, e^{Kh} + e^{\overline{K}j}) \in F$, for an arbitrary $h \in L$, imply that $\mu^T e^{\overline{K}j} = \mu_{\overline{K}j} = 0$. Note that the second point belongs to F even if $j \in L \cup \{i\}$;
- $\mu_{Kj} = 0$ and $\pi_j = 0$ for all $j \in V \setminus (L \cup \{i\})$: We apply induction on d_j, the distance from j to the star $L \cup \{i\}$ in G, that is, $d_j = \min\{\delta(j,v) \mid v \in L \cup \{i\}\}$. For $d \geq 1$, let $V_d = \{j \in V \setminus (L \cup \{i\}) : d_j \leq d\}$. If $j \in V_1$ and $(j,h) \in E(G)$ for some $h \in L$, we use the points $(e, e^{Kh}) \in F$ and $(e, e^{Kh} + e^{Kj}) \in F$ to get $\mu_{Kj} = 0$. If $j \in V_1$ and $(j,h) \notin E(G)$ for all vertex $h \in L$, then $(j,i) \in E(G)$ and we use the points $(e, \sum_{u \in H[L \cup \{i,j\}]} e^{Ku}) \in F$ and $(e, \sum_{u \in H[L \cup \{i\}]} e^{Ku}) \in F$ to get $\mu_{Kj} + \sum_{w \in H[L \cup \{i,j\}] \setminus H[L \cup \{i\}]} \mu_{Kw} = 0$. Since it was proved that $\mu_{Kw} = 0$ for every $w \in V \setminus (L \cup \{i\})$ with $(w,h) \in E(G)$ for some $h \in L$, then $\sum_{w \in H[L \cup \{i,j\}] \setminus H[L \cup \{i\}]} \mu_{Kw} = 0$, which leads to $\mu_{Kj} = 0$.

  Suppose that $\mu_{Kj} = 0$ for all $j \in V_d$, for some $d \geq 1$. Now, consider $j \in V_{d+1}$. If its distance is determined by a vertex $h \in L$, then the point $(e, \sum_{u \in H[\{h,j\}]} e^{Ku})$ belongs to F. Since $H[\{h,j\}] \setminus \{h,j\} \subseteq V_d$, we have that $\mu_{Kw} = 0$, for all $w \in H[\{h,j\}] \setminus \{h,j\}$. Using such a point and $(e, e^{Kh}) \in F$, we conclude that $\mu^T e^{Kj} = \mu_{Kj} = 0$. Otherwise, if its distance is determined by i, the point $(e, \sum_{u \in H[L \cup \{i,j\}]} e^{Ku})$ belongs to F. Since $H[L \cup \{i,j\}] \setminus \{i,j\} \subseteq V_d$, we have that $\mu_{Kw} = 0$, for all $w \in H[L \cup \{i,j\}] \setminus \{i,j\}$. Using such a point and $(e, \sum_{u \in H[L \cup \{i\}]} e^{Ku}) \in F$, we conclude that $\mu^T e^{Kj} = \mu_{Kj} = 0$.

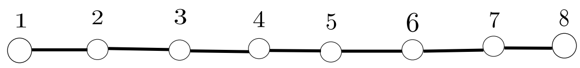
  Similarly, we can prove that $\pi_j = 0$ for all $j \in V \setminus (L \cup \{i\})$;
- $\pi_i = 0$: If $K = K(i)$, the points $(e, \sum_{u \in H[L \cup \{i\}]} e^{Ku}) \in F$ and $(e - e^i, \sum_{u \in H[L \cup \{i\}]} e^{Ku}) \in F$ lead to $\pi_i = 0$. Otherwise, we choose the points $(e, e^{Kh}) \in F$, for some $h \in L$, and

$(e - e^i, e^{Kh} + e^{\bar{K}i}) \in F$ to get $\pi_i = 0$, since $\mu_{\bar{K}i}$ was proved to be zero;

- $\pi_h = 0$ for all $h \in L$: If $K = K(j)$, the points $(e, e^{Kh}) \in F$ and $(e - e^h, e^{Kh}) \in F$ lead to $\pi_j = 0$. Otherwise, the points $(e, e^{Kh'}) \in F$ and $(e - e^h, e^{Kh'} + e^{\bar{K}h}) \in F$, for some $h' \in L \setminus \{h\}$, imply $\pi_h + \mu_{\bar{K}h} = 0$. Since $\mu_{\bar{K}h}$ was proved to be zero previously, we get $\pi_j = 0$;
- $\mu_{Kh} = \mu_{Kh'}$ for all $h, h' \in L$, $h \neq h'$: The points $(e, e^{Kh}) \in F$ and $(e, e^{Kh'}) \in F$ show that $\mu_{Kh} = \mu_{Kh'}$;
- $\mu_{Ki} = -(|L| - 1)\mu_{Kh}$ for any $h \in L$: The points $(e, e^{Kh'}) \in F$, for some $h' \in L$, and $(e, \sum_{u \in H[L \cup \{i\}]} e^{Ku}) \in F$ imply that $\mu_{Ki} + \sum_{h \in L \setminus \{h'\}} \mu_{Kh} + \sum\limits_{w \in H[L \cup \{i\}] \setminus (L \cup \{i\})} \mu_{Kw} = 0$. Since $\mu_{Kw} = 0$ for all $w \in H[L \cup \{i\}] \setminus (L \cup \{i\})$, we get $\mu_{Ki} = -(|L| - 1)\mu_{Kh'}$. ■

In (LIMA, 2011), a generalization of the convexity inequalities (4.17) was presented for the *Path Convex Recoloring Problem* (*PCR*). We noted that these generalized inequalities are also valid for the geodesic classification problem. They are counterparts of the alternating path inequalities (4.12) and will be called *generalized convexity inequalities*. An example can be seen in Figure 19.

Figure 19 – An example of generalized convexity inequality for ILP2.



$$z_{Kv1} - z_{Kv3} + z_{Kv4} - z_{Kv6} + z_{Kv8} \leq 1, \quad K \in \{B, R\}$$

Proposition 4.2.30 *Let $S_{hj} =< h = l_1, q_1, \ldots, l_t, q_t, j >$, $t \leq \lfloor \delta(h, j) \rfloor$, be a sequence of distinct vertices with odd cardinality that corresponds to an incomplete or complete shortest path from h to j in G. So, the following inequalities are valid for P_2:*

$$(\sum_{i=1}^{t} z_{Kl_i}) + z_{Kj} - (\sum_{i=1}^{t} z_{Kq_i}) \leq 1, \qquad K \in \{B, R\}. \tag{4.21}$$

Proof Note that each triple $\phi = (h, j, t)$ can induce one or more inequalities for a given class K. We say that such an inequality is induced by $\phi = (h, j, t)$.

Analogous to the Proposition 4.2.13, we are going to prove that the generalized convexity inequalities are valid for P_2 using induction on t. Since when $t = 1$, the corresponding inequalities are exactly the convexity inequalities (4.17) of ILP2. Let $K \in \{B,R\}$ and consider the sum of the following three valid inequalities for $t = 1$:

$$\begin{array}{rrrrrcl}
+z_{Kl_1} & -z_{Kq_1} & +z_{Kl_2} & & & \leq & 1 \\
 & & +z_{Kl_2} & -z_{Kq_2} & +z_{Kj} & \leq & 1 \\
+z_{Kl_1} & & & -z_{Kq_2} & +z_{Kj} & \leq & 1 \\
\hline
2z_{Kl_1} & -z_{Kq_1} & +2z_{Kl_2} & -2z_{Kq_2} & +2z_{Kj} & \leq & 3.
\end{array}$$

Thus, we have that inequality $2z_{Kl_1} - z_{Kq_1} + 2z_{Kl_2} - 2z_{Kq_2} + 2z_{Kj} \leq 3$ is valid for P_2. Summing this inequality with the inequality $-z_{Kq_1} \leq 0$, we get:

$2z_{Kl_1} - 2z_{Kq_1} + 2z_{Kl_2} - 2z_{Kq_2} + 2z_{Kj} \leq 3.$

Dividing the inequality above by 2 yields:

$z_{Kl_1} - z_{Kq_1} + z_{Kl_2} - z_{Kq_2} + z_{Kj} \leq \dfrac{3}{2}.$

Since we want integer valued solutions for the variables z, we can round down the right-hand side to get the following stronger valid inequality for P_2:

$z_{Kl_1} - z_{Kq_1} + z_{Kl_2} - z_{Kq_2} + z_{Kj} \leq 1.$

Therefore, all generalized convexity inequalities $\phi = (h, j, 2)$ are valid for P_2. We can generalize this procedure for the inequalities induced by $\phi(h, j, t)$, $t \geq 3$. To do so, we sum up the following four inequalities:

- the inequality induced by $\phi(h = l_1, l_t, t-1)$;
- the inequality induced by $\phi(l_2, j, t-1)$;
- the inequality $z_{Kl_1} - z_{Kq_t} + z_{Kj} \leq 1$;
- the inequality $-z_{Kq_1} \leq 0$.

After that, we divide by 2 the resulting inequality and round down the right-hand side. ■

We now show that the generalized convexity inequalities are facet-defining when the base sequence corresponds to a complete shortest path.

Proposition 4.2.31 *Let $S_{hj} =< h = l_1, q_1, \ldots, l_t, q_t, j >$ be a shortest path from h to j in G. So, the following inequalities are facet-defining for P_2:*

$$(\sum_{i=1}^{t} z_{Kl_i}) + z_{Kj} - (\sum_{i=1}^{t} z_{Kq_i}) \leq 1, \qquad K \in \{B,R\}. \tag{4.22}$$

Proof Let $K \in \{B,R\}$ and $F = \{(o,z) \in P_2 : (\sum_{i=1}^{t} z_{Kl_i}) + z_{Kj} - (\sum_{i=1}^{t} z_{Kq_i}) = 1\}$. Assume that $F \subseteq F' := \{(o,z) \in P_2 : \pi^T o + \mu^T z = \pi_0\}$. We prove that $\pi^T o + \mu^T z = \pi_0$ is a multiple of $(\sum_i z_{Kl_i}) + z_{Kj} - (\sum_i z_{Kq_i}) = 1$ by cases, as follows (for this proof, we consider $l_{t+1} = j$):

- $\mu_{\overline{K}\ell} = 0$ for all $\ell \in V$: The points $(e, e^{Kh}) \in F$ and $(e, e^{Kh} + e^{\overline{K}\ell}) \in F$ imply that $\mu^T e^{\overline{K}\ell} = \mu_{\overline{K}\ell} = 0$. Note that the second point belongs to F even if $\ell = h$;
- $\mu_{K\ell} = 0$ and $\pi_\ell = 0$ for all $\ell \in V \setminus V(S_{hj})$: We apply induction on d_ℓ, the distance from ℓ to the path S_{hj} in G, that is, $d_\ell = \min\{\delta(\ell, v) \mid v \in V(S_{hj})\}$. For $d \geq 1$, let $V_d = \{\ell \in V \setminus V(S_{hj}) : d_\ell \leq d\}$. If $\ell \in V_1$ and $(\ell, l_i) \in E(G)$, for some $i \in \{1, \ldots, t+1\}$, we use the points $(e, e^{Kl_i}) \in F$ and $(e, e^{Kl_i} + e^{K\ell}) \in F$ to get $\mu_{K\ell} = 0$. If $\ell \in V_1$ and $(\ell, l_i) \notin E(G)$ for all $i \in \{1, \ldots, t+1\}$, then $(\ell, q_i) \in E(G)$, for some $i \in \{1, \ldots, t\}$, and we use the points $(e, \sum_{u \in H[\{l_i, q_i, l_{i+1}\}]} e^{Ku}) \in F$ and $(e, \sum_{u \in H[\{l_i, q_i, l_{i+1}, \ell\}]} e^{Ku}) \in F$ to get $\mu_{K\ell} + \sum_{w \in H[\{l_i, q_i, l_{i+1}, \ell\}] \setminus H[\{l_i, q_i, l_{i+1}\}]} \mu_{Kw} = 0$. Since $\mu_{Kw} = 0$ for every $w \in V \setminus V(S_{hj})$ with $(w, l_j) \in E(G)$ for some $j \in \{1, \ldots, t+1\}$, then $\sum_{w \in H[\{l_i, q_i, l_{i+1}, \ell\}] \setminus H[\{l_i, q_i, l_{i+1}\}]} \mu_{Kw} = 0$, which leads to $\mu_{K\ell} = 0$.

  Suppose that $\mu_{K\ell} = 0$ for all $\ell \in V_d$, for some $d \geq 1$. Now, consider $\ell \in V_{d+1}$. If its distance is determined from a vertex l_i, $i \in \{1, \ldots, t+1\}$, then the point $(e, \sum_{u \in H[\{l_i, \ell\}]} e^{Ku})$ belongs to F. Since $H[\{l_i, \ell\}] \setminus \{l_i, \ell\} \subseteq V_d$, we have that $\mu_{Kw} = 0$, for all $w \in H[\{l_i, \ell\}] \setminus \{l_i, \ell\}$. Using such a point and $(e, e^{Kl_i}) \in F$, we conclude that $\mu^T e^{K\ell} = \mu_{K\ell} = 0$. Otherwise, if its distance is determined from q_i, $i \in \{1, \ldots, t\}$, the point $(e, \sum_{u \in H[\{l_i, q_i, l_{i+1}, \ell\}]} e^{Ku})$ belongs to F. Since $H[\{l_i, q_i, l_{i+1}, \ell\}] \setminus \{l_i, q_i, l_{i+1}, \ell\} \subseteq V_d$, we have that $\mu_{Kw} = 0$, for all $w \in H[\{l_i, q_i, l_{i+1}, \ell\}] \setminus \{l_i, q_i, l_{i+1}, \ell\}$. Using such a point and $(e, \sum_{u \in H[\{l_i, q_i, l_{i+1}\}]} e^{Ku}) \in F$, we conclude that $\mu^T e^{K\ell} = \mu_{K\ell} = 0$.

  Similarly, we can prove that $\pi_\ell = 0$ for all $\ell \in V \setminus (V(S_{hj}))$;
- $\pi_{l_i} = 0$ for all $i \in \{1, \ldots, t+1\}$: If $K = K(l_i)$, the points $(e, e^{Kl_i}) \in F$ and $(e - e^{l_i}, e^{Kl_i}) \in F$ lead to $\pi_i = 0$. Otherwise, we choose the points $(e, e^{Kl_u}) \in F$, for some $u \in \{1, \ldots, t+1\} \setminus$

$\{i\}$, and $(e-e^{l_i}, e^{Kl_u}+e^{\bar{K}l_i}) \in F$ to get $\pi_i = 0$, since $\mu_{\bar{K}l_i}$ was proved to be zero;

- $\pi_{q_i} = 0$ for all $i \in \{1,\ldots,t\}$: If $K = K(q_i)$, the points $(e, \sum_{u \in H[\{l_i,q_i,l_{i+1}\}]} e^{Ku}) \in F$ and $(e - e^{q_i}, \sum_{u \in H[\{l_i,q_i,l_{i+1}\}]} e^{Ku}) \in F$ lead to $\pi_{q_i} = 0$. Otherwise, we choose the points $(e, e^{Kl_u}) \in F$, for some $u \in \{1,\ldots,t+1\}$, and $(e-e^{q_i}, e^{Kl_u}+e^{\bar{K}q_i}) \in F$ to get $\pi_i = 0$, since $\mu_{\bar{K}q_i}$ was proved to be zero previously;
- $\mu_{Kl_i} = \mu_{Kl_{i'}}$ for all $i,i' \in \{1,\ldots,t+1\}$, $i \neq i'$: The points $(e, e^{Kl_i}) \in F$ and $(e, e^{Kl_{i'}}) \in F$ show that $\mu_{Kl_i} = \mu_{Kl_{i'}}$;
- $\mu_{Kq_i} = -\mu_{Kl_i}$ for all $i \in \{1,\ldots,t\}$: The points $(e, \sum_{u \in H[\{l_i,q_i,l_{i+1}\}]} e^{Ku}) \in F$ and $(e, e^{Kl_{i+1}}) \in F$ imply that $\mu_{Kl_i} + \mu_{Kq_i} + \sum_{w \in H[\{l_i,q_i,l_{i+1}\}] \setminus \{l_i,q_i,l_{i+1}\}} \mu_{Kw} = 0$. Since $\mu_{Kw} = 0$ for all $w \in H[\{l_i,q_i,l_{i+1}\}] \setminus \{l_i,q_i,l_{i+1}\}$, we get $\mu_{Kq_i} = -\mu_{Kl_i}$. ■

The corollary below follows from Proposition 4.2.31.

Corollary 4.2.32 *Let $h,j,i \in V$ such that $(h,i) \in E(G)$, $(i,j) \in E(G)$ and $(h,j) \notin E(G)$. Then, constraint* (4.17) *defines a facet of P_2.*

A generalization of inequalities (4.21) can be obtained by allowing some vertices to appear more than once in the base sequence. In such cases, it relates to walks, not paths. The following proposition shows this generalization and its validity.

Proposition 4.2.33 *Let $S =< v_1, v_2, \ldots, v_{2t+1} >$, $t \geq 1$, be a sequence of vertices with odd cardinality that corresponds to an incomplete or complete walk in G such that*

$$\{v_{2i}, \ldots, v_{2j}\} \cap D_{v_{2i-1}v_{2j+1}} \neq \emptyset, \qquad \forall 1 \leq i \leq j \leq t. \tag{4.23}$$

Then, the inequalities below are valid for P_2:

$$\sum_{i=0}^{t} z_{Kv_{2i+1}} - \sum_{i=1}^{t} z_{Kv_{2i}} \leq 1, \qquad K \in \{B,R\}. \tag{4.24}$$

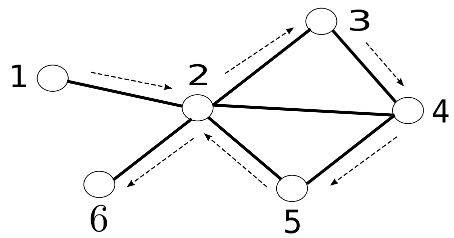
Proof Let $K \in \{B,R\}$ and suppose that (4.24) is not valid for P_2, that is, there is an integer point $(o,z) \in P_2$ such that $\sum_{i=0}^{t} z_{Kv_{2i+1}} - \sum_{i=1}^{t} z_{Kv_{2i}} \geq 2$. Then, there are i,j, such that $1 \leq i \leq j \leq t$,

$z_{Kv_{2i-1}} = z_{Kv_{2j+1}} = 1$ and $z_{Kv_l} = 0\ \forall l \in [2i, 2j]$. Particularly, $z_{Kv_{2l}} = 0\ \forall l \in [i, j]$, which contradicts (4.23) and (4.17). ■

The requirements (4.23) mean that, for any two vertices i, j in odd positions in the sequence S, there is at least one vertex w in an even position in S between them such that $w \in D_{ij}$. As a consequence, a given vertex v can not appear in two different odd positions in S because $D_{vv} = \emptyset$. However, the same vertex can appear many times in even positions.

We call inequalities (4.24) *generalized walk inequalities*. An example is shown in Figure 20. Note that $S = \{v_1, v_2, v_3, v_4, v_5, v_2, v_6\}$, $v_2 \in D_{v_1v_3} \cap D_{v_1v_5} \cap D_{v_1v_6} \cap D_{v_3v_6} \cap D_{v_5v_6}$ and $v_4 \in D_{v_3v_5}$, which satisfy (4.23).

Figure 20 – Generalized walk inequality example.



$$S =< v_1, v_2, v_3, v_4, v_5, v_2, v_6 >$$

$$z_{Kv1} + z_{Kv3} + z_{Kv5} + z_{Kv6} - 2z_{Kv2} - z_{Kv4} \leq 1, \quad K \in \{B, R\}$$

As a counterpart of the generalized C_4 inequalities (4.10) given for P_1, we now present valid inequalities for P_2 that also involve variables z for vertices in V_N.

Proposition 4.2.34 *Let $S_B = \{i, i'\}$ and $S_B = \{j, j'\}$ be subsets of $V(G)$ such that $S_B \subseteq V_{BN}$, $S_R \subseteq V_{RN}$. If $j \in H[\{i, i'\}]$, $j' \in H[\{i, i'\}]$, $i \in H[\{j, j'\}]$ and $i' \in H[\{j, j'\}]$, thus the following inequality is valid for P_2:*

$$\sum_{i \in S_B \cap V_N} (1 - z_{Bi}) + \sum_{i \in S_B \cap V_B} o_i + \sum_{i \in S_R \cap V_N} (1 - z_{Ri}) + \sum_{i \in S_R \cap V_R} o_i \geq 2. \tag{4.25}$$

If the subgraph induced by $S_B \cup S_R$ is exactly a C_4, then (4.25) are facet-defining, as stated below.

Proposition 4.2.35 *Let $S_B = \{i, i'\}$ and $S_R = \{j, j'\}$ be subsets of $V(G)$ such that $S_B \subseteq V_{BN}$, $S_R \subseteq V_{RN}$. If $S_B \cup S_R$ is an induced C_4, thus the following inequality is a facet-defining for P_2:*

$$\sum_{i \in S_B \cap V_N} (1 - z_{Bi}) + \sum_{i \in S_B \cap V_B} o_i + \sum_{i \in S_R \cap V_N} (1 - z_{Ri}) + \sum_{i \in S_R \cap V_R} o_i \geq 2. \tag{4.26}$$

Proof By following the same approach used in the proof of Propositions 4.2.29 and 4.2.31, we can prove by induction that $\sum_{i \in S_B \cap V_N} (1 - z_{Bi}) + \sum_{i \in S_B \cap V_B} o_i + \sum_{i \in S_R \cap V_N} (1 - z_{Ri}) + \sum_{i \in S_R \cap V_R} o_i \geq 2$ defines a facet of P_2. ■

5 PIECEWISE LINEAR SEPARATION FOR THE GEODESIC CLASSIFICATION PROBLEM

Another way of dealing with linear inseparability (instead of just considering outliers) is to divide the blue and red sets of samples in subsets, which are called *groups*, and then impose the convexity constraints for each group separately, as shown in Figure 22. In this example, the set of blue samples is divided into two groups, A_{B_1} and A_{B_2}, whereas the set of red samples forms a single group A_R. Note that each pair of blue and red groups is linearly separable in the sense of Definition 4.0.1.

Figure 21 – An example of instance for the geodesic classification problem.

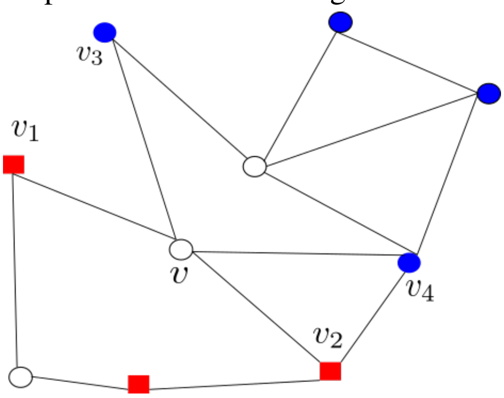
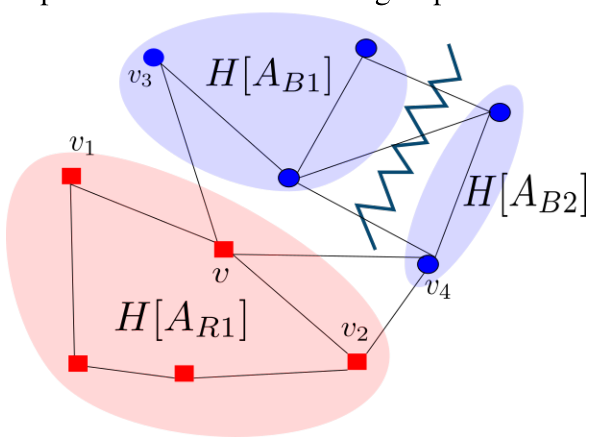


Figure 22 – An example of solution with multi-group for the example of Figure 21.



As in 2-class multi-group Euclidean classification (Problem 2), we can think of an approach to geodesic convexity classification that combines two strategies: identification of outliers and division of the non-outlier vertices in groups. As a result, we require any pair of opposite color groups to be linearly separable and aims at minimizing the number of outliers. To formalize this idea, we introduce the following definition.

Definition 5.0.1 *Let $F_B \subseteq 2^{V_B}$ and $F_R \subseteq 2^{V_R}$ be families of subsets of V_B and V_R, respectively, and $A_N \subseteq V_N$. A triple (F_B, F_R, A_N) is* piecewise linearly separable (with respect to G) *if*

(M1) $H[A] \cap A' = \emptyset$, $\quad A \in F_B$, $A' \in F_R$,

(M2) $H[A'] \cap A = \emptyset$, $\quad A \in F_B$, $A' \in F_R$, *and*

(M3) $H[A] \cap H[A'] \cap A_N = \emptyset$, $\quad A \in F_B$, $A' \in F_R$,

and piecewise linearly inseparable *otherwise, where $F_B = \{A \mid A \text{ is a blue group}\}$ (resp., $F_R = \{A \mid A \text{ is a red group}\}$). We can omit the word* piecewise *when it can be understood by the context. For the sake of simplicity, we refer to (F_B, F_R, A_N) simply as (F_B, F_R) if, and only if, $A_N = V_N$.*

To define the geodesic classification problem with multi-groups, we use two fundamental parameters, L_B and L_R, that represent an upper bound on the number of groups, for each class respectively, that can be used in a solution. Let $L_{BR} = L_B + L_R$. We denote by C_B (resp., C_R) the set of indices of the blue (resp., red) groups, $C_{BR} = C_B \cup C_R$, $V_{BN} = V_B \cup V_N$ and $V_{RN} = V_R \cup V_N$. Thus, $L_B = |C_B|$ and $L_R = |C_R|$. When we use $K \in \{B, R\}$ to specify a class, $\bar{K}$ denotes the opposite class. Besides, $K(i)$ and $\bar{K}(i)$ denote the color of the initially classified vertex $i \in V_{BR}$ and its opposite color, respectively. The definition of the problem is described below.

Problem 4. 2-*class Multi-group Geodesic Classification Problem* (2-*MGC*):

Given a connected simple graph $G = (V, E)$, sets of initially classified vertices V_B (blue vertices) and V_R (red vertices), with $V_N = V \setminus (V_{BR})$, and upper bound parameters L_B, L_R, find groups $A_k \subseteq V_B$, $\forall k \in \{1, \dots, L_B\}$, and $A_{k'} \subset V_R$, $\forall k' \in \{1, \dots, L_R\}$, such that:

(M0) $(\bigcup_{k \in C_B} A_k, \bigcup_{k' \in C_R} A_{k'})$ satisfies (M1), (M2) and (M3),

(M4) $A_k \cap A_j = \emptyset$, $\quad \forall k, j \in C_B$ or $\forall k, j \in C_R$, and

(M5) $|V_{BR}| - |(\bigcup_{k \in C_B} A_k \cup \bigcup_{k' \in C_R} A_{k'})|$ is minimum.

The vertices in $V_{BR} \setminus \bigcup_{k \in C_{BR}} A_k$ are the outliers. It is worth observing that this problem reduces to the 2-class single-group geodesic classification problem when $|L_B| = |L_R| = 1$.

In the solution of the 2-MGC problem, G is divided in blue and red convex sets given by $H[A_k]$, $k \in C_B$, and $H[A_{k'}]$, $k' \in C_R$, respectively. Vertices in V_N that belong to a blue (resp., red) convex set are set to the blue (resp., red) class. In addition, there can be vertices in V_N that do not belong to any convex set. It is worth remarking that the classification of such vertices is not in the scope of this work. For a vertex *i* and a convex set *k*, we define *i* as active if *i* belongs to *k*, and inactive otherwise.

Note that there is no constraint about intersection of convex sets of the same class, that is $H[A_k] \cap H[A_{k'}]$, $k,k' \in C_B$ (or $k,k' \in C_R$). However, convex sets related to opposite classes can intercept only at outliers. Other important point to note is that condition (M4) could be removed without changing the optimum. Indeed, consider a solution satisfying (M1)-(M3). Suppose $K \in \{B,R\}$ and $l \in A_k \cap A_{k'}$, for some $k,k' \in C_K$, $k \neq k'$. Define $\hat{A}_j = A_j, \forall j \in C_K$, $j \neq k$, and $\hat{A}_k = A_k \backslash \{l\}$. Thus, $\hat{A}_j \subseteq A_j$ and $H[\hat{A}_j] \subseteq H[A_j]$ and $\hat{A}_j \subseteq A_j$, $\forall j \in C_K$, which implies that conditions (M1), (M2) and (M3) remain satisfied. Also, $(\bigcup_{j \neq k} A_j) \cup A_k = (\bigcup_{j \neq k} \hat{A}_j) \cup \hat{A}_k$, keeping the same number of outliers and, consequently, the same value for the objective function. By repeating this process, we initially get a solution satisfying (M4).

5.1 Computational complexity of the 2-MGC problem

Let L_o be a parameter corresponding to the maximum number of outliers in a solution of the 2-MGC problem. We denote by 2-*MGCD*(G,V_B,V_R,L_B,L_R,L_o) the following decision problem:

Problem 5. 2-*class Multi-group Geodesic Classification Decision Problem* (2-*MGCD*):

Given a connected simple graph $G = (V,E)$, sets of initially classified vertices V_B (blue vertices) and V_R (red vertices), with $V_N = V \backslash (V_{BR})$, and upper bound parameters L_B, L_R, L_o, are there groups $A_k \subseteq V_B$, $k \in \{1,\ldots,L_B\}$, and $A_{k'} \subset V_R$, $k' \in \{1,\ldots,L_R\}$, such that:

1. $(\bigcup_{k \in C_B} A_k, \bigcup_{k' \in C_R} A_{k'})$ satisfies (M1), (M2) and (M3),
2. $A_k \cap A_j = \emptyset$, $\quad \forall k,j \in C_B$ or $\forall k,j \in C_R$, and
3. $|V_{BR}| - |(\bigcup_{k \in C_B} A_k \cup \bigcup_{k' \in C_R} A_{k'})| \leq L_o$?

We conjecture that this decision problem is NP-complete even for $|L_B| = |L_R| = 1$. Let us observe its similarity with the problems studied in (ARTIGAS *et al.*, 2011) and (BUZATU; CATARANCIUC, 2015), which were proved to be NP-complete problems. In the first reference, the authors proved that deciding whether a graph can be partitioned into 2 (disjoint) convex sets is NP-complete. On the other hand, the same complexity was proved in the second reference for the covering by 2 convex sets. The 2-MGCD problem looks for “improper"covering/packings in the sense that some vertices in V_{BR} and some vertices in V_N may not be covered by some convex subsets whereas some vertices in V_{BR} (up to L_o) may be in the intersection of the convex subsets.

We have shown, in Subsection 4.1.1, that this problem for $L_B = L_R = 1$ can be view as a special case of the well-known set covering problem. Actually, in the next subsection, we obtain a similar result for the general case. Although we have not established its computational complexity yet, we know some cases where it can be solved in polynomial time. They are described below:

1. $L_B \geq |V_B|$ and $L_R \geq |V_R|$: a trivial solution can be obtained by setting each vertex to its own group;
2. $|V_{BR}| \leq 2$ or $|V_B| = |V|$ or $|V_R| = |V|$: a trivial solution can be obtained defining one group with all blue vertices and another group with all red vertices;
3. $L_o \geq \min\{|V_B|, |V_R|\}$: a trivial solution can be obtained by defining each vertex in V_B (or V_R, if $|V_R| < |V_B|$) as an outlier (it gives an upper bound for the problem);
4. L_o or V_{BR} is constant: we can use a brute force algorithm;
5. Path instances with fixed L_B and L_R: For this type of instances, we can find an optimal solution using a brute force algorithm. It suffices to divide the path into $L_B + L_R$ parts, where some of them may be empty, and to choose the class of each part. By combinatorial calculation, the number of such possibilities is less than or equal to $\binom{2(L_B+L_R-1)+|V_{BR}|-1}{L_B+L_R-1} \cdot \binom{L_B+L_R}{L_B}$, which is polynomial.

5.2 Integer formulations for the 2-MGC problem

In this section, we present the analog of those formulations presented in Section 4. In addition, a third formulation is obtained by a variable transformation from ILP2.

5.2.1 A set covering formulation for the 2-MGC problem

To define the analogous formulation to ILP1, we now include a new index k in the variables, which represents a group. Thus, for each $i \in V_{BR}$ and $k \in C_{K(i)}$,

$$o_{ki} = \begin{cases} 1, & \text{if } i \in V_{BR} \setminus A_k, \\ 0, & \text{if } i \in A_k. \end{cases}$$

Then, the formulation becomes:

$$(\text{ILP1}^M) \ \min \sum_{i \in V_{BR}} \sum_{k \in C_{K(i)}} o_{ki} - \sum_{K \in \{B,R\}} |V_K|(L_K - 1) \tag{5.1}$$

$$\text{st:} \sum_{k \in C_{K(i)}} o_{ki} \geq L_{K(i)} - 1, \qquad \forall i \in V_{BR}, \tag{5.2}$$

$$\sum_{j \in S} o_{\bar{k}j} + o_{ki} \geq 1, \qquad \begin{array}{c} \forall i \in V_{BR}, \forall k \in C_{K(i)}, \forall \bar{k} \in C_{\bar{K}(i)}, \\ \forall S \subseteq V_{\bar{K}(i)} : i \in H[S], \end{array} \tag{5.3}$$

$$\sum_{j \in S} o_{kj} + \sum_{j \in T} o_{\bar{k}j} \geq 1, \qquad \begin{array}{c} \forall k \in C_B, \forall \bar{k} \in C_R, \\ \forall S \subseteq V_B \, \forall T \subseteq V_R : H[S] \cap H[T] \cap V_N \neq \emptyset, \end{array} \tag{5.4}$$

$$o \in \mathbb{B}^{|V_B|L_B + |V_R|L_R}. \tag{5.5}$$

Constraints (5.2) indicate that each vertex in V_{BR} can not belong to more than one group (condition (M4)). The slackness of (5.2) indicates whether i is an outlier or not. It is important to note that this condition is related to groups but not to their convex hull. Constraints (5.3) ensure that, if an initially classified vertex belongs to the convex hull of a group of its opposite class, then it must be an outlier, i.e., it can not belong to any group of its class (conditions (M1) and (M2)). The V_N-free intersection of convex sets of opposite classes is described by (5.4) (condition (M3)). If $i \in V_N$ belongs to the convex hull of a blue (resp. red) class, then i is assigned to the blue (resp. red) class. Finally, by summing up the slackness of (5.2), we see that the objective function minimizes the number of outliers (condition (M5)).

Below, we translate the informal argument presented above into a formal proof of the correctness of the formulation.

Proposition 5.2.1 *Formulation ILP*1^M *is correct.*

Proof First, observe that conditions (M1), (M2) and (M3) are a replication of conditions (C1), (C2) and (C3) for each pair $(k,\bar{k}) \in C_B \times C_R$. Similarly, for each pair $(k,\bar{k}) \in C_B \times C_R$, constraints (5.3)-(5.4) are exactly constraints (4.1)-(4.2), which model conditions (C1), (C2) and

(C3), according to Proposition (4.1.1). Besides, constraints (5.2) model condition (M4). Indeed, they are equivalent to $\sum_{k\in C_{K(i)}}(1-o_{ki}) \leq 1$, for all $i \in V_{BR}$, which ensure that each vertex $i \in V_{BR}$ takes part in at most one set A_k, $k \in C_{K(i)}$. Finally, observe that, due to (5.2), the objective function counts exactly the number of vertices from V_{BR} outside $\bigcup_{k\in C_{BR}} A_k$. ■

We remark that constraints (5.2) could be discarded from ILP1^M, as condition (M4) is not necessary for the definition of 2-MGC. However, because of them, we can replace (5.3) by a smaller set of strengthened constraints, as follows.

Proposition 5.2.2 *Let* $o \in \mathbb{B}^{|V_B|L_B+|V_R|L_R}$ *satisfying* (5.2). *Then, o satisfies* (5.3) *if, and only if, o satisfies*

$$\sum_{j\in S} o_{\bar{k}j} + \sum_{k\in C_{K(i)}} o_{ki} \geq L_{K(i)}, \quad \forall i \in V_{BR}, \forall \bar{k} \in C_{\bar{K}(i)}, \forall S \subseteq V_{\bar{K}(i)} : i \in H[S]. \tag{5.6}$$

Proof First, suppose that o satisfies all constraints in (5.3). Let $i \in V_{BR}$, $\bar{k} \in C_{\bar{K}(i)}$ and $S \subseteq V_{\bar{K}(i)}$ such that $i \in H[S]$. If $o_{ki} = 1$ for all $k \in C_{K(i)}$, then (5.6) trivially holds. Otherwise, by (5.3), $\sum_{j\in S} o_{\bar{k}j} \geq 1$. This together with (5.2) implies (5.6).

Now suppose that o does not satisfy one of the constraints in (5.3). Then, there are $i \in V_{BR}$, $k \in C_{K(i)}$, $\bar{k} \in C_{\bar{K}(i)}$ and $S \subseteq V_{\bar{K}(i)}$ with $i \in H[S]$ such that $\sum_{j\in S} o_{\bar{k}j} = 0$ and $o_{ki} = 0$. Therefore, the left-hand side of (5.6) becomes $\sum_{c\in C_{K(i)\setminus\{k\}}} o_{ci}$, which is at most $L_{K(i)} - 1$. It follows that one constraint in (5.6) is not satisfied. ■

5.2.2 *The associated polytope -* P_1^M

Let P_1^M be the polytope associated with ILP1^M, that is,

$$P_1^M = \text{conv}\{o \in \mathbb{B}^{|V_B|L_B+|V_R|L_R} : (5.2)-(5.4)\}.$$

Recall that constraints (5.2) could be discarded from ILP1^M. Thus, we also consider the relaxed polytope:

$$\bar{P}_1^M = \text{conv}\{o \in \mathbb{B}^{|V_B|L_B+|V_R|L_R} : (5.3)-(5.4)\}.$$

It is worth observing that P_1^M and $\bar{P}_1^M$ are equal to P_1 when $L_B = L_R = 1$.

Proposition 5.2.3 *P_1^M and $\bar{P}_1^M$ are full-dimensional.*

Proof Consider the $|V_B|L_B + |V_R|L_R + 1$ points e and $e - e^{ki}$ for each $i \in V_{BR}$ and $k \in C_{K(i)}$. Since they are affinely independent and belong to P_1^M, this set is full-dimensional. Besides, $\bar{P}_1^M \supseteq P_1^M$ implies that $\bar{P}_1^M$ is full-dimensional. ■

Similarly to Proposition 4.1.4, we can deduce that any facet-defining inequality different from a bounding inequality has non-negative coefficients.

Proposition 5.2.4 *Let $\pi^T o \geq \pi_0$ be a facet-defining inequality of P_1^M or $\bar{P}_1^M$. If it is different from $o_{ki} \leq 1$ and $o_{ki} \geq 0$, for all $i \in V_{BR}$ and $k \in C_{K(i)}$, then $\pi \geq 0$ and $\pi_0 > 0$.*

5.2.3 Relations with P_1

In this subsection, we relate polytope P_1^M with P_1 through affine transformations. These relations will yield the conversion of valid inequalities for one polytope into valid inequalities for the other one.

Let us partition any vector $o \in \mathbb{B}^{|V_{BR}|}$ into (o^B, o^R), where o^B and o^R comprise the components indexed by $i \in V_B$ and $i \in V_R$, respectively. For the ease of the notation, we write $o = (o^B, o^R)$. Similarly, for $o \in \mathbb{B}^{|V_B|L_B + |V_R|L_R}$ and $k \in C_{BR}$, let o^k comprise the components o_{ki}, for all $i \in V_{BR}$ such that $K(i) = k$. Thus, we may highlight the components of o related to $k \in C_{BR}$ and write $o = (o^k, o')$, where o' comprises the remaining components.

We start by defining an injective affine transformation that maps P_1 to $P_1^M \subseteq \bar{P}_1^M$.

Proposition 5.2.5 *Let $k \in C_B$ and $\bar{k} \in C_R$. Let $\mathscr{Q}^{k\bar{k}} : \mathbb{R}^{|V_{BR}|} \to \mathbb{R}^{|V_B|L_B + |V_R|L_R}$ be the affine transformation such that $\mathscr{Q}^{k\bar{k}}(o^B, o^R) = (o^k, o^{\bar{k}}, o')$, where*

$$\begin{bmatrix} o^k \\ o^{\bar{k}} \\ o' \end{bmatrix} = \begin{bmatrix} I_{|V_B|} & 0 \\ 0 & I_{|V_R|} \\ 0 & 0 \end{bmatrix} \begin{bmatrix} o^B \\ o^R \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ e \end{bmatrix}. \tag{5.7}$$

Then, $\mathscr{Q}^{k\bar{k}}(P_1) \subseteq P_1^M$.

Proof It suffices to prove that an integer point $(o^B, o^R) \in P_1$ is mapped to an integer point $(o^k, o^{\bar{k}}, o') \in P_1^M$. Since $o^{k'} = e$ for all $k' \in C_{BR} \setminus \{k, \bar{k}\}$, Inequalities (5.2) are satisfied. It also implies that all inequalities (5.3)-(5.4), expect for those related to the pair $(k, \bar{k})$, are trivially satisfied. Besides, each of those inequalities associated with $(k, \bar{k})$ holds, provided that (o^B, o^R) satisfies (4.1)-(4.2). ■

It is worth remarking that the inclusion stated in Proposition 5.2.5 can be strict. Moreover, it can occur that $\bigcup_{(k,\bar{k}) \in C_B \times C_R} \mathscr{Q}^{k\bar{k}}(P_1) \subsetneq P_1^M$. For example, consider $G = (V_B \cup V_R \cup V_N, E)$ as a path with 3 vertices, where $V_B = \{v_1, v_3\}$ contains the end vertices, $V_R = \{v_2\}$ contains the middle vertex and $V_N = \emptyset$. Let $C_B = \{b, b'\}$ and $C_R = \{r\}$. Observe that the point $o \in \mathbb{B}^5$ where the non-null components are $o_{bv_1} = o_{b'v_3} = 1$ belongs to P_1^M. However, $\mathscr{Q}^{br}(P_1)$ yields either $o_{b'v_1} = o_{b'v_3} = 1$ whereas $\mathscr{Q}^{b'r}(P_1)$ yields either $o_{bv_1} = o_{bv_3} = 1$, implying that $P_1^M \supsetneq \mathscr{Q}^{br}(P_1) \cup \mathscr{Q}^{b'r}(P_1)$.

In the following two lemmas we identify points in P_1^M or $\bar{P}_1^M$ that are not directly mapped by these affine transformations from points in P_1. Despite this, such transformations are still useful in this context. For ease of the notation, we use $\mathscr{Q}^{k\bar{k}}$ and $\mathscr{Q}^{\bar{k}k}$ indistinctly (provided that $k \in C_B$ and $\bar{k} \in C_R$, or $k \in C_R$ and $\bar{k} \in C_B$).

Lemma 5.2.6 *Let $i \in V_{BR}$, $k \in C_{K(i)}$ and $\bar{k} \in C_{\bar{K}(i)}$. If $o \in P_1$ satisfies $\sum_{j \in S} o_j \geq 1$, for all $S \subseteq V_{\bar{K}(i)}$ such that $i \in H[S]$, then $\mathscr{Q}^{k\bar{k}}(o) - e^{k'i} \in \bar{P}_1^M$, for all $k' \in C_{K(i)} \setminus \{k\}$.*

Proof Let $o \in P_1$ with $\sum_{j \in S} o_j \geq 1$, for all $S \subseteq V_{\bar{K}(i)}$ such that $i \in H[S]$. To obtain the claimed result, it suffices to assume that o is integer and show that $\hat{o} = \mathscr{Q}^{k\bar{k}}(o) - e^{k'i}$ satisfies (5.3)-(5.4). Observe that

$$\hat{o}_{cj} = \begin{cases} o_j, & \text{if } c \in \{k, \bar{k}\}, \\ 0, & \text{if } c = k' \text{ and } j = i, \\ 1, & \text{otherwise.} \end{cases} \tag{5.8}$$

By contradiction, first suppose that $\hat{o}$ violates (5.3) for some $\ell \in V_{BR}$, $c \in C_{K(\ell)}$, $\bar{c} \in C_{\bar{K}(\ell)}$ and $S \subseteq V_{\bar{K}(\ell)}$ such that $\ell \in H[S]$. Then, $\sum_{j \in S} \hat{o}_{\bar{c}j} = \hat{o}_{c\ell} = 0$. Observe that we must have $\{c, \bar{c}\} \neq \{k, \bar{k}\}$; otherwise, by (4.1) we would have $\sum_{j \in S} \hat{o}_{\bar{c}j} + \hat{o}_{c\ell} = \sum_{j \in S} o_j + o_\ell \geq 1$. On the other hand, since $|S| \geq 2$ and $\sum_{j \in S} \hat{o}_{\bar{c}j} = 0$, it must be $\bar{c} \in \{k, \bar{k}\}$, and so $c \notin \{k, \bar{k}\}$. As $\hat{o}_{c\ell} = 0$, it follows that $c = k'$ and $\ell = i$. Thus, $S \subseteq V_{\bar{K}(i)}$ and $i \in H[S]$. Therefore, by hypothesis, we have $\sum_{j\in S} \hat{o}_{\bar{c}j} = \sum_{j\in S} o_j \geq 1$: a contradiction.

Now suppose that $\hat{o}$ violates (5.4) for some $c \in C_B$, $\bar{c} \in C_R$, $S \subseteq V_B$ and $T \subseteq V_R$ such that $H[S] \cap H[T] \cap V_N \neq \emptyset$. This means that $\sum_{j\in S} \hat{o}_{cj} = \sum_{j\in T} \hat{o}_{\bar{c}j} = 0$. Since $|S| \geq 2$ and $|T| \geq 2$, it must be $\{c, \bar{c}\} = \{k, \bar{k}\}$. Then, by (4.2), $\sum_{j\in S} \hat{o}_{cj} + \sum_{j\in T} \hat{o}_{\bar{c}j} = \sum_{j\in S\cup T} o_j \geq 1$: a contradiction. ■

Corollary 5.2.7 *Let $i \in V_{BR}$, $k \in C_{K(i)}$ and $\bar{k} \in C_{\bar{K}(i)}$. If $o \in P_1$ is such that $o_i = 0$, then $\mathscr{Q}^{k\bar{k}}(o) - e^{k'i} \in \bar{P}_1^M$, for all $k' \in C_{K(i)} \setminus \{k\}$.*

Proof By Constraints (4.1), $o_i = 0$ implies $\sum_{j\in S} o_j \geq 1$, for all $S \subseteq V_{\bar{K}(i)}$ such that $i \in H[S]$. The result then follows by Lemma 5.2.6. ■

Lemma 5.2.8 *Let $i \in V_{BR}$, $k \in C_{K(i)}$ and $\bar{k} \in C_{\bar{K}(i)}$. Let $o \in P_1$ satisfying $o_i = 1$ and $\sum_{j\in S} o_j \geq 1$, for all $S \subseteq V_{\bar{K}(i)}$ such that $i \in H[S]$, then $\mathscr{Q}^{k\bar{k}}(o) - e^{k'i} \in P_1^M$, for all $k' \in C_{K(i)} \setminus \{k\}$.*

Proof Let $o \in P_1$ with $o_i = 1$ and $\sum_{j\in S} o_j \geq 1$, for all $S \subseteq V_{\bar{K}(i)}$ such that $i \in H[S]$. It suffices to prove the result when o is integer. By Lemma 5.2.6, $\hat{o} = \mathscr{Q}^{k\bar{k}}(o) - e^{k'i}$ satisfies (5.3)-(5.4). Then, it remains to consider Constraints (5.2). Recall the expression of $\hat{o}$ in (5.8). Let $\ell \in V_{BR}$. If $\ell \neq i$, then $\sum_{c\in C_{K(\ell)}} \hat{o}_{c\ell} = o_\ell + |C_{K(\ell)}| - 1 \geq L_{K(\ell)} - 1$. If $\ell = i$, then $\sum_{c\in C_{K(\ell)}} \hat{o}_{c\ell} = o_i + |C_{K(\ell)}| - 2 = L_{K(\ell)} - 1$. In both cases, $\hat{o}$ satisfies (5.2). ■

We can now show that P_1 is a projection of P_1^M and $\bar{P}_1^M$.

Corollary 5.2.9 *Let $k \in C_B$ and $\bar{k} \in C_R$. Then, $P_1 = \mathrm{proj}_{(o^k, o^{\bar{k}})} P_1^M = \mathrm{proj}_{(o^k, o^{\bar{k}})} \bar{P}_1^M$.*

Proof By Proposition 5.2.5, we conclude that $P_1 \subseteq \mathrm{proj}_{(o^k, o^{\bar{k}})} P_1^M$. Besides, $\mathrm{proj}_{(o^k, o^{\bar{k}})} P_1^M \subseteq \mathrm{proj}_{(o^k, o^{\bar{k}})} \bar{P}_1^M$ due to $P_1^M \subseteq \bar{P}_1^M$. Now, let $o = (o^B, o^R) \in \mathrm{proj}_{(o^k, o^{\bar{k}})} \bar{P}_1^M$. Then, for all $i \in V_{BR}$ and $S \subseteq V_{\bar{K}(i)}$ such that $i \in H[S]$, Constraints (5.3) lead to $\sum_{j\in S} o_j + o_i \geq 1$. Similarly, for all $S \subseteq V_B$ and $T \subseteq V_R$ such that $H[S] \cap H[T] \cap V_N \neq \emptyset$, Constraints (5.4) imply that $\sum_{j\in S} o_j + \sum_{j\in T} o_j \geq 1$. Then, o satisfies (4.1)-(4.2). In particular, we have shown that the integer points in $\mathrm{proj}_{(o^k, o^{\bar{k}})} \bar{P}_1^M$

belong to P_1. It follows that $\text{proj}_{(o^k,o^{\bar{k}})}\bar{P}_1^M \subseteq P_1$. Therefore, $P_1 = \text{proj}_{(o^k,o^{\bar{k}})}P_1^M = \text{proj}_{(o^k,o^{\bar{k}})}\bar{P}_1^M$. ■

A surjective affine mapping from a subset of P_1^M to P_1 can also be established.

Proposition 5.2.10 *Let $\mathscr{Q}$ be any of the affine transformations defined in Proposition 5.2.5 and $\mathscr{R} : \mathbb{R}^{|V_B|L_B+|V_R|L_R} \to \mathbb{R}^{|V_{BR}|}$ be the affine transformation such that $\mathscr{R}(o) = \tilde{o}$, where*

$$\tilde{o}_i = \sum_{k\in C_{K(i)}} o_{ki} - L_{K(i)} + 1 \quad \forall i \in V_{BR}. \tag{5.9}$$

Then $\mathscr{R}(\mathscr{Q}(o^{BR})) = o^{BR}$, for all $o^{BR} = (o^B, o^R) \in \mathbb{R}^{|V_{BR}|}$, and $P_1 \subseteq \mathscr{R}(P_1^M)$.

Proof Let $o^{BR} = (o^B, o^R) \in \mathbb{R}^{|V_{BR}|}$, $\hat{o} = \mathscr{Q}(o^{BR}) = (o^B, o^R, e)$, and $\tilde{o} = \mathscr{R}(\hat{o})$. Let $i \in V_K$. Assume that $\mathscr{Q} = \mathscr{Q}^{k\bar{k}}$, for $k \in C_K$, $\bar{k} \in C_{\bar{K}}$ and $K \in \{B,R\}$. Then, $\tilde{o}_i = \sum_{k'\in C_{K(i)}} \hat{o}_{k'i} - L_{K(i)} + 1 = \hat{o}_{ki} = o_i^{BR}$, that is, $\mathscr{R}(\mathscr{Q}(o^{BR})) = o^{BR}$. In particular, $P_1 = \mathscr{R}(\mathscr{Q}(P_1))$. By Proposition 5.2.5, it follows that $P_1 \subseteq \mathscr{R}(P_1^M)$. ■

The example presented after Proposition 5.2.5 also shows that it can occur the strict inclusion $P_1 \subsetneq \mathscr{R}(P_1^M)$. Indeed, for the point $o \in B^5$ given in that example, we have $\mathscr{R}(o) = 0$, but $0 \notin P_1$. Anyway, Proposition 5.2.10 ensures that valid inequalities for $\mathscr{R}(P_1^M)$ are valid for P_1. In other terms, we can derive valid inequalities for P_1 from valid inequalities for P_1^M, as follows.

Proposition 5.2.11 *If $\sum_{k\in C_{BR}}(\pi^k)^T o^k \geq \pi_0$ is valid for P_1^M then $(\pi^k)^T o^B + (\pi^{\bar{k}})^T o^R \geq \pi_0 - \sum_{k'\in C_{BR}\setminus\{k,\bar{k}\}}(\pi^{k'})^T e$ is valid for P_1, for every $k \in C_B$ and $\bar{k} \in C_R$.*

Proof Let $k \in C_B$ and $\bar{k} \in C_R$. Let $\mathscr{Q}^{k\bar{k}}(o) = Qo + q$ be defined as in Proposition 5.2.5, with matrix Q and vector q given by (5.7). Then, $\mathscr{Q}^{k\bar{k}}(P_1) \subseteq P_1^M$, and so $\sum_{k\in C_{BR}}(\pi^k)^T o^k \geq \pi_0$ is valid for $\mathscr{Q}^{k\bar{k}}(P_1)$. By Proposition 2.2.4, it follows that $(\pi^k)^T o^B + (\pi^{\bar{k}})^T o^R \geq \pi_0 - \sum_{k'\in C_{BR}\setminus\{k,\bar{k}\}}(\pi^k)^T e$ is valid for P_1. ■

We can also obtain a valid inequality for P_1 by combining coefficients, from a valid inequality for P_1^M, related to different groups of the same color.

Proposition 5.2.12 *If* $\sum_{k \in C_{BR}} (\pi^k)^T o^k \geq \pi_0$ *is valid for* P_1^M, *then* $\sum_{i \in V_{BR}} \pi_{k_i i} o_i \geq \pi_0 - \sum_{i \in V_{BR}} \sum_{k \in C_{K(i)} \setminus \{k_i\}} \pi_{ki}$ *is valid for* P_1, *where* $\pi_{k_i i} = \min\{\pi_{ki} : k \in C_{K(i)}\}$ *for all* $i \in V_{BR}$.

Proof Let $\tilde{o} \in P_1$. By Proposition 5.2.10, there is $o \in P_1^M$ such that $\tilde{o} = \mathscr{R}(o)$, that is, $\tilde{o}_i = \sum_{k \in C_{K(i)}} o_{ki} - L_{K(i)} + 1 = \sum_{k \in C_{K(i)}} (o_{ki} - 1) + 1$ for all $i \in V_{BR}$. It follows that

$$\begin{aligned}
\sum_{i \in V_{BR}} \pi_{k_i i} \tilde{o}_i &= \sum_{i \in V_{BR}} \sum_{k \in C_{K(i)}} \pi_{k_i i}(o_{ki} - 1) + \sum_{i \in V_{BR}} \pi_{k_i i} \\
&\geq \sum_{i \in V_{BR}} \sum_{k \in C_{K(i)}} \pi_{ki}(o_{ki} - 1) + \sum_{i \in V_{BR}} \pi_{k_i i} \qquad \text{(due to } \pi_{k_i i} \leq \pi_{ki}, o_{ki} \leq 1\text{)} \\
&\geq \pi_0 + \sum_{i \in V_{BR}} \sum_{k \in C_{K(i)} \setminus \{k_i\}} \pi_{ki} \qquad \text{(due to } \sum_{k \in C_{BR}} (\pi^k)^T o^k \geq \pi_0\text{)}.
\end{aligned}$$

Therefore, $\sum_{i \in V_{BR}} \pi_{k_i i} o_i \geq \pi_0 - \sum_{i \in V_{BR}} \sum_{k \in C_{K(i)} \setminus \{k_i\}} \pi_{ki}$ is valid for P_1. ■

Note that $\pi_{k_i i}$ is the "best" coefficient for o_i in a $\geq$-inequality among the coefficients of o_{ki}, for $k \in C_{K(i)}$. However, the decrease in the right-hand side needed to make the inequality with these best coefficients valid leads this inequality to be dominated by the inequalities defined in Proposition 5.2.11. Indeed, given $k \in C_B$ and $\bar{k} \in C_R$, the summation of $(\pi^k)^T o^B + (\pi^{\bar{k}})^T o^R \geq \pi_0 - \sum_{k' \in C_{BR} \setminus \{k, \bar{k}\}} (\pi^k)^T e$ with the valid inequalities $(\pi_{ki} - \pi_{k_i i}) o_i \leq (\pi_{ki} - \pi_{k_i i})$, for all $i \in C_B$, and $(\pi_{\bar{k}i} - \pi_{k_i i}) o_i \leq (\pi_{\bar{k}i} - \pi_{k_i i})$, for all $i \in C_R$, gives exactly $\sum_{i \in V_{BR}} \pi_{k_i i} o_i \geq \pi_0 - \sum_{i \in V_{BR}} \sum_{k' \in C_{K(i)} \setminus \{k_i\}} \pi_{k' i}$.

5.2.4 Valid inequalities and facets

Corollary 5.2.9 and Proposition 2.2.10 allow us to directly derive valid inequalities for $\bar{P}_1^M$ (and consequently for P_1^M) from those valid for P_1.

Proposition 5.2.13 *If* $\pi^T o^B + \lambda^T o^R \geq \pi_0$ *is valid for* P_1 *then* $\pi^T o^k + \lambda^T o^{\bar{k}} \geq \pi_0$ *is valid for* P_1^M *and* $\bar{P}_1^M$, *for all* $k \in C_B$ *and* $\bar{k} \in C_R$.

In the following, we focus on facetness properties. For easiness of the presentation, we start considering the upper bounding constraints.

Proposition 5.2.14 *For every $i \in V_{BR}$ and $k \in C_{K(i)}$, $o_{ki} \leq 1$ defines a facet of P_1^M and $\bar{P}_1^M$.*

Proof Consider the points e and $e - e^{k'j}$, for each $j \in V_{BR}$ and $k' \in C_{K(j)}$ with $(k', j) \neq (k,i)$. They are affinely independent points of P_1^M in the face defined by $o_{ki} \leq 1$. Then, this face is actually a facet of P_1^M. The same occurs for $\bar{P}_1^M \supseteq P_1^M$. ■

We can show that $\bar{P}_1^M$ inherits the facetness properties that hold for P_1.

Proposition 5.2.15 *If $\pi^T o^B + \lambda^T o^R \geq \pi_0$ is facet-defining for P_1 then $\pi^T o^k + \lambda^T o^{\bar{k}} \geq \pi_0$ is facet-defining for $\bar{P}_1^M$, for all $k \in C_B$ and $\bar{k} \in C_R$.*

Proof Let $k \in C_B$ and $\bar{k} \in C_R$. By Proposition 5.2.14, it remains to consider the case where $\pi^T o^B + \lambda^T o^R \geq \pi_0$ is different from $o_i \leq 1$, for all $i \in V_{BR}$. Let $F = \{(o^B, o^R) \in P_1 : \pi^T o^B + \lambda^T o^R = \pi_0\}$ and $F^M = \{(o^k, o^{\bar{k}}, o') \in \bar{P}_1^M : \pi^T o^k + \lambda^T o^{\bar{k}} = \pi_0\}$. By Proposition 5.2.13, F^M is a face of $\bar{P}_1^M$. Consider the following points:

1. $\mathscr{Q}^{k\bar{k}}(O)$, where O is a set of $p = |V_B| + |V_R|$ affinely independent points of F. By Proposition 5.2.5, $\mathscr{Q}^{k\bar{k}}(O) \subseteq \bar{P}_1^M$. Besides, every point $(o^k, o^{\bar{k}}, o') \in \mathscr{Q}^{k\bar{k}}(o^B, o^R)$ has $o^k = o^B$, $o^{\bar{k}} = o^R$ and $o' = e$. Therefore, we can deduce that $\mathscr{Q}^{k\bar{k}}(O)$ comprises p affinely independent points in F^M.
2. For every $i \in V_{BR}$ and $k' \in C_{K(i)} \setminus \{k, \bar{k}\}$, point $\mathscr{Q}^{k\bar{k}}(o) - e^{k'i}$, where o is a solution in F with $o_i = 0$ (such a solution exists because $F \neq \{o \in P_1 : o_i = 1\}$). Let $\hat{O}$ be the set formed by these points. Then, $\hat{O} \subset \bar{P}_1^M$ by Corollary 5.2.7. Besides, every point $(o^k, o^{\bar{k}}, o') = \mathscr{Q}^{k\bar{k}}(o^B, o^R) - e^{k'i}$ has $o^k = o^B$, $o^{\bar{k}} = o^R$ and $o' = e - e^{k'i}$. It follows that $\hat{O}$ has $|V_B|(L_B - 1) + |V_R|(L_R - 1)$ affinely independent points in F^M.

Since the o'-components in $\mathscr{Q}^{k\bar{k}}(O)$ are always 1 and each point in $\hat{O}$ has zero in one of these components, we can conclude that the whole set of points is affinely independent. Therefore, as $O \cup \hat{O} \subseteq F^M$, we conclude that F^M is a facet of $\bar{P}_1^M$. ■

By applying Proposition 5.2.15 to the facet-defining inequalities presented in Subsections 4.2.1 and 4.2.2, we get facet-defining inequalities of $\bar{P}_1^M$, as follows.

Corollary 5.2.16 *The following inequalities define facets of $\bar{P}_1^M$.*

1. $o_{ki} \geq 0$, $\forall i \in V_{BR}$, $k \in C_{K(i)}$,
2. *The multi-group V_{BR}-disjoint $\mathscr{N}$-set elementary inequalities (corresponding to* (4.4)*, which includes the corresponding generalized* 3*-path inequalities* (4.7)*) satisfying the conditions of Theorem 4.2.2:* $\sum_{j\in S} o_{\bar{k}j} + o_{ki} \geq 1$*, for every* $k \in C_{K(i)}$ *and* $\bar{k} \in C_{\bar{K}(i)}$*,*
3. *The multi-group V_N-disjoint $\mathscr{N}$-set elementary inequalities (corresponding to* (4.5)*, which includes the corresponding X-swing inequalities* (4.8)*) satisfying the conditions of Theorem 4.2.2:* $\sum_{j\in S} o_{kj} + \sum_{j\in T} o_{\bar{k}j} \geq 1$*, for every* $k \in C_B$ *and* $\bar{k} \in C_R$*,*
4. *The multi-group generalized C_4 inequality (corresponding to* (4.10)*):* $o_{ki} + o_{ki'} + o_{\bar{k}j} + o_{\bar{k}j'} \geq 2$*, for every* $k \in C_B$*,* $\bar{k} \in C_R$*,*
5. *The multi-group $\mathscr{N}$-set inequalities (corresponding to* (4.6)*, which includes the corresponding star tree* (4.11) *and alternating path* (4.12) *inequalities) satisfying the conditions of Theorem 4.2.2:* $\sum_{i\in S} \frac{o_{ki}}{v_i} + \sum_{j\in T} \frac{o_{\bar{k}j}}{v_j} \geq 1$*, for every* $k \in C_B$*,* $\bar{k} \in C_R$*.*

Regarding P_1^M, the presence of constraints (5.2) may require additional conditions to obtain facet-defining inequalities from P_1.

Proposition 5.2.17 *Let $\pi^T o^B + \lambda^T o^R \geq \pi_0$ be a valid inequality for P_1 defining a facet F such that, for each $i \in V_{BR}$ with $L_{K(i)} > 1$, there is $o \in F$ such that $o_i = 1$ and $\sum_{j\in S} o_j \geq 1$ for every $S \subseteq V_{\bar{K}(i)}$ with $i \in H[S]$. Then $\pi^T o^k + \lambda^T o^{\bar{k}} \geq \pi_0$ is facet-defining for P_1^M, for all $k \in C_B$ and $\bar{k} \in C_R$.*

Proof Let $k \in C_B$, $\bar{k} \in C_R$ and $F^M = \{(o^k, o^{\bar{k}}, o') \in P_1^M : \pi^T o^k + \lambda^T o^{\bar{k}} = \pi_0\}$. By Proposition 5.2.13, F^M is a face of P_1^M. Consider the following points in F^M obtained from points in $F = \{(o^B, o^R) \in P_1 : \pi^T o^B + \lambda^T o^R = \pi_0\}$:

1. $\mathscr{Q}^{k\bar{k}}(O)$, where O is a set of $p = |V_B| + |V_R|$ affinely independent points of F. By Proposition 5.2.5, $\mathscr{Q}^{k\bar{k}}(O) \subseteq P_1^M$. Besides, every point $(o^k, o^{\bar{k}}, o') \in \mathscr{Q}^{k\bar{k}}(o^B, o^R)$ has $o^k = o^B$, $o^{\bar{k}} = o^R$ and $o' = e$. This implies that $\mathscr{Q}^{k\bar{k}}(O)$ comprises p affinely independent points in F^M.
2. For every $i \in V_{BR}$ and $k' \in C_{K(i)} \setminus \{k, \bar{k}\}$, point $\mathscr{Q}^{k\bar{k}}(o) - e^{k'i}$, where o is a solution in F with $o_i = 1$ and $\sum_{j\in S} o_j \geq 1$ for all $S \subseteq V_{\bar{K}(i)}$ such that $i \in H[S]$ (this solution exists by the hypotheses on $\pi^T o^B + \lambda^T o^R \geq \pi_0$; this case only exists when $L_{K(i)} > 1$). Let $\hat{O}$ be the set formed by these points. Then, $\hat{O} \subset P_1^M$ by Lemma 5.2.8. Besides, every point

$(o^k, o^{\bar{k}}, o') = \mathscr{Q}^{k\bar{k}}(o^B, o^R) - e^{k'i}$ has $o^k = o^B$, $o^{\bar{k}} = o^R$ and $o' = e - e^{k'i}$. Therefore, $\hat{O}$ has $|V_B|(L_B - 1) + |V_R|(L_R - 1)$ affinely independent points in F^M.

Since the o'-components in $\mathscr{Q}^{k\bar{k}}(O)$ are always 1 and each point in $\hat{O}$ has zero in one of these components, we can conclude that the whole set of points is affinely independent. Therefore, as $O \cup \hat{O} \subseteq F^M$, we conclude that F^M is a facet of P_1^M. ■

Due to the condition stated in Proposition 5.2.17, only some of the inequalities enumerated in Corollary 5.2.16 are also facet-defining for P_1^M.

Corollary 5.2.18 *The following inequalities define facets of P_1^M.*

1. *The multi-group V_N-disjoint $\mathscr{N}$-set elementary inequalities (corresponding to* (4.5)*, which includes the corresponding X-swing inequalities* (4.8)*) satisfying the conditions of Theorem 4.2.2: $\sum_{j \in S} o_{kj} + \sum_{j \in T} o_{\bar{k}j} \geq 1$, for every $k \in C_B$ and $\bar{k} \in C_R$,*
2. *The multi-group generalized C_4 inequalities* (4.10)*: $o_{ki} + o_{ki'} + o_{\bar{k}j} + o_{\bar{k}j'} \geq 2$, for every $k \in C_B$, $\bar{k} \in C_R$.*

It is worth remarking that the conditions stated in Proposition 5.2.17 do not hold for the non-negativity constraints or constraints (4.1) of ILP1. In this case, the corresponding constraints in ILP1^M are dominated by (5.2) and (5.6), respectively. These stronger inequalities are actually facet-defining for P_1^M. To obtain these and other facets, we use the following proposition.

Proposition 5.2.19 *Let $\sum_{j \in S} o_j + \pi_i o_i \geq \pi_0$, $\pi_0 \geq 0$, be facet-defining for P_1, where $i \in V_{BR}$ and $S \subseteq V_{\bar{K}(i)}$ with $i \in H[S]$. Let $\bar{k} \in C_{\bar{K}(i)}$. The following inequality is facet-defining for P_1^M:*

$$\sum_{j \in S} o_{\bar{k}j} + \pi_i \sum_{k \in C_{K(i)}} o_{ki} \geq \pi_0 + \pi_i (L_{K(i)} - 1). \tag{5.10}$$

Proof First, note that $\tilde{o}_j = 1$, for all $j \in V_{K(i)}$, and $\tilde{o}_j = 0$, for all $j \in V_{\bar{K}(i)}$, defines a point in P_1. Since $\sum_{j \in S} \tilde{o}_j + \pi_i \tilde{o}_i \geq \pi_0$, it follows that $\pi_i \geq \pi_0$.

To show validity of (5.10), let o be an integer point in P_1^M. If $\sum_{k \in C_{K(i)}} o_{ki} = L_{K(i)}$, the inequality holds since $\pi_i \geq \pi_0$. Otherwise, we have $\sum_{k \in C_{K(i)}} o_{ki} = L_{K(i)} - 1$ by (5.2), which means that $o_{k'i} = 0$ for some $k' \in C_{K(i)}$. Now let $\tilde{o} \in \mathbb{B}^{|V_{BR}|}$ be such that $\tilde{o}_j = o_{k'j}$, for all $j \in V_{K(i)}$, and $o_j = o_{\bar{k}j}$, for all $j \in V_{\bar{K}(i)}$. Since o satisfies (5.3)-(5.4), we have that $\tilde{o}$ satisfies (4.1)-(4.2). This means that $\tilde{o} \in P_1$, leading to $\pi_0 \leq \sum_{j\in S} \tilde{o}_j + \pi_i \tilde{o}_i = \sum_{j\in S} o_{\bar{k}j} + \pi_i o_{k'i} = \sum_{j\in S} o_{\bar{k}j}$. Therefore, $\sum_{j\in S} o_{\bar{k}j} + \pi_i \sum_{k\in C_{K(i)}} o_{ki} \geq \pi_0 + \pi_i(L_{K(i)} - 1)$.

Now, we show facetness. Let F^M be the face of P_1^M defined by (5.10). Let k be an arbitrary element of $C_{K(i)}$. Consider the following $|V_B|L_B + |V_R|L_R$ points:

1. $\mathscr{Q}^{k\bar{k}}(O)$, where $O \subseteq P_1$ is set of $|V_{BR}|$ affinely independent integer points satisfying $\sum_{j\in S} o_j + \pi_i o_i \geq \pi_0$ at equality. These points do exist due to the hypothesis. By Proposition 5.2.5, $\mathscr{Q}^{k\bar{k}}(O) \subseteq P_1^M$. Recall that every point $(o^k, o^{\bar{k}}, o') \in \mathscr{Q}^{k\bar{k}}(O)$ has $o' = e$, $o_{cj} = o_j$, for all $j \in V_{BR}$ and $c \in C_{K(j)} \cap \{k, \bar{k}\}$. This implies that $\mathscr{Q}^{k\bar{k}}(O)$ comprises $|V_{BR}|$ affinely independent points in P_1^M. Moreover,

$$\sum_{j\in S} o_{\bar{k}j} + \pi_i \sum_{k'\in C_{K(i)}} o_{k'i} = \sum_{j\in S} o_j + \pi_i(o_i + L_{K(i)} - 1)$$
$$= \sum_{j\in S} o_j + \pi_i o_i + \pi_i(L_{K(i)} - 1) = \pi_0 + \pi_i(L_{K(i)} - 1).$$

Thus, $\mathscr{Q}^{k\bar{k}}(O) \subseteq F^M$.

2. $\hat{O} = \{\mathscr{Q}^{k\bar{k}}(\bar{o}) - e^{c'\ell} + e^{c\ell} \mid \ell \in V_{BR}, c' \in C_{K(\ell)} \setminus \{k, \bar{k}\}, \{c\} = C_{K(\ell)} \cap \{k, \bar{k}\}\}$, where $\bar{o}$ is an arbitrary point with $\bar{o}_\ell = 0$ in the facet of P_1 induced by $\sum_{j\in S} o_j + \pi_i o_i \geq \pi_0$ (such a point do exist because this inequality is different from $o_\ell \leq 1$ as $\pi_0 \geq 0$). Let $\hat{o} = \mathscr{Q}^{k\bar{k}}(\bar{o}) - e^{c'\ell} + e^{c\ell} \in \hat{O}$. By item 1, $\tilde{o} = \mathscr{Q}^{k\bar{k}}(\bar{o}) \in F^M$. Since we are adding $e^{c\ell}$ and subtracting $e^{c'\ell}$ to $\tilde{o}$ and $c, c' \in C_{K(\ell)}$, $\hat{o} \in \hat{O}$ still satisfies (5.2) and (5.6) (and so (5.3)). Moreover, recall that, for all $j \in V_{BR}$ and $k' \in C_{K(j)}$, we have $\tilde{o}_{k'j} = \bar{o}_j$, if $k' \in \{k, \bar{k}\}$, and $\tilde{o}_{k'j} = 1$, if $k' \notin \{k, \bar{k}\}$. In particular, $\tilde{o}_{c\ell} = \bar{o}_\ell = 0$ and $\tilde{o}_{c'\ell} = 1$. Therefore, $\hat{o}$ is a binary point. Besides, since $\hat{o}_{c'j} = \tilde{o}_{c'j} = 1$ for all $j \neq \ell$, $\tilde{o}$ satisfies constraints (5.4), and $|S| \geq 2$ and $|T| \geq 2$ in these constraints, we conclude that $\hat{o}$ satisfies (5.4). It follows that $\hat{o} \in P_1^M$. Furthermore, since $\tilde{o} \in F^M$ satisfies $\sum_{j\in S} o_{\bar{k}j} + \pi_i \sum_{k\in C_{K(i)}} o_{ki} \geq \pi_0 + \pi_i(L_{K(i)} - 1)$ at equality, the same happens with $\hat{o} = \tilde{o} - e^{c'\ell} + e^{c\ell}$. Therefore, $\hat{O} \subseteq F^M$. Let us also remark that $\hat{O}$ is an affinely independent set with $|V_B|(L_B - 1) + |V_R|(L_R - 1)$ points.

Finally, consider a point $(o^k, o^{\bar{k}}, o') \in \mathscr{Q}^{k\bar{k}}(O) \cup \hat{O}$. Since the o'-components of any point in $\mathscr{Q}^{k\bar{k}}(O)$ are all 1 and exactly one of them turns to be zero in each point of $\hat{O}$ (a different component for each point), we conclude that the entire set of points is affinely independent. Therefore, F^M is a facet of P_1^M. ■

By applying Proposition 5.2.19 to the facet-defining inequalities of P_1 considered in Proposition 4.2.4, Corollary 4.2.3, Corollary 4.2.7 and Proposition 4.2.12, we obtain the following facet-defining inequalities.

Corollary 5.2.20 *The following inequalities define facets of P_1^M.*

1. *Constraints* (5.2) *(corresponding to $o_i \geq 0$),*
2. *The multi-group generalized* 3*-path inequalities (corresponding to* (4.7)*) satisfying the conditions of Corollary 4.2.7: $o_{\bar{k}h} + o_{\bar{k}j} + \sum_{k \in C_{K(i)}} o_{ki} \geq L_{K(i)}$, for every $\bar{k} \in C_{barK(i)}$,*
3. *Constraints* (5.6) *that satisfy the conditions of Corollary 4.2.3,*
4. *The multi-group star tree inequalities (corresponding to* (4.11)*) satisfying the conditions of Theorem 4.2.2: $\sum_{j \in L} o_{\bar{k}j} + (|L|-1) \sum_{k \in C_{K(i)}} o_{ki} \geq (|L|-1) L_{K(i)}$, for every $\bar{k} \in C_{\bar{K}(i)}$.*

As a consequence of Corollary 5.2.20)(1), we can still establish the following result.

Corollary 5.2.21 *Let $i \in V_{BR}$ and $k \in C_{K(i)}$. Then $o_{ki} \geq 0$ defines a facet of P_1^M if, and only, if $L_{K(i)} = 1$.*

Proof If $L_{K(i)} = 1$, $o_{ki} \geq 0$ is the same as (5.2), which is facet-defining, according to Corollary 5.2.20(1). If $L_{K(i)} \geq 2$, $o_{ki} \geq 0$ is the summation of (5.2) and $-o_{k'i} \geq -1$, for all $k' \in C_{K(i)} \setminus \{i\}$. ■

To end this section, we relate ILP1^M with the formulation that we introduced in (ARAÚJO *et al.*, 2019). Although presented for the single-group case, it can be readily generalized for the multi-group case, as follows:

$$\min \sum_{i \in V_{BR}} \sum_{k \in C_{K(i)}} a_{ki} \tag{5.11}$$

$$\text{st: } \sum_{k \in C_{K(i)}} a_{ki} \leq 1, \qquad \forall i \in V_{BR}, \tag{5.12}$$

$$\sum_{j \in S} a_{\bar{k}j} + a_{ki} \leq |S|, \qquad \begin{array}{c} \forall i \in V_{BR}, \forall k \in C_{K(i)}, \forall \bar{k} \in C_{\bar{K}(i)}, \\ \forall S \subseteq V_{\bar{K}(i)} : i \in H[S], \end{array} \tag{5.13}$$

$$\sum_{j \in S} a_{kj} + p_i \leq |S|, \qquad \begin{array}{c} \forall i \in V_N, \forall k \in C_B, \\ \forall S \subseteq V_B : i \in H[S], \end{array} \tag{5.14}$$

$$\sum_{j \in T} a_{\bar{k}j} + (1 - p_i) \leq |T|, \qquad \begin{array}{c} \forall i \in V_N, \forall \bar{k} \in C_R, \\ \forall T \subseteq V_R : i \in H[T], \end{array} \tag{5.15}$$

$$a \in \mathbb{B}^{|V_B| L_B + |V_R| L_R}, \tag{5.16}$$

$$p \in \mathbb{B}^{|V_N|}. \tag{5.17}$$

The variables a_{ki} indicate whether a vertex i is chosen to form a group (basis) A_k. Thus, we have $a_{ki} = 1 - o_{ki}$. With this variable transformation, we can see that (5.11) is equivalent to the objective function (5.1) whereas constraints (5.12) and (5.13) are direct rewritten of (5.2) and (5.3). In addition, constraints (5.14) and (5.15) are a disaggregation of (5.4) by means of a binary variable p_i, $i \in V_N$, to indicate whether i is reached by a blue convex set ($p_i = 0$) or a red convex set ($p_i = 1$). This guarantees that condition (M3) holds. Observe that (5.4) is the summation of (5.14) and (5.15).

As one can infer from the relation stated above, formulation (5.11)-(5.17) is essentially ILP1^M (with the addition of variable p_i, for all $i \in V_N$). For this reason, we do not consider it in this text.

5.2.5 A compact formulation for the 2-MGC problem

Similarly to the definition of ILP1^M, we add an index k to the o variables in ILP2 in order to represent groups. Thus, we have for each $i \in V_{BR}$ and $k \in C_{K(i)}$,

$$o_{ki} = \begin{cases} 1, & \text{if } i \in V_{BR} \setminus A_k, \\ 0, & \text{if } i \in A_k, \end{cases}$$

and, for each $k \in C_{BR}$ and $i \in V$,

$$z_{ki} = \begin{cases} 1, & \text{if } i \in H[A_k], \\ 0, & \text{otherwise.} \end{cases}$$

The z variables have exactly the same meaning as in ILP2. The formulation is the following.

$$(\text{ILP2}^M)\ \min \sum_{i \in V_{BR}} \sum_{k \in C_{K(i)}} o_{ki} - \sum_{K \in \{B,R\}} |V_K|(L_K - 1) \tag{5.18}$$

$$\text{st:} \sum_{k \in C_{K(i)}} o_{ki} \geq L_{K(i)} - 1, \qquad \forall i \in V_{BR}, \tag{5.19}$$

$$o_{ki} \geq z_{\bar{k}i}, \qquad \forall i \in V_{BR}, k \in C_{K(i)}, \bar{k} \in C_{\bar{K}(i)}, \tag{5.20}$$

$$z_{ki} + z_{\bar{k}i} \leq 1, \qquad \forall i \in V_N, k \in C_B, \bar{k} \in C_R, \tag{5.21}$$

$$z_{ki} + o_{ki} \geq 1, \qquad \forall i \in V_{BR}, k \in C_{K(i)}, \tag{5.22}$$

$$z_{kh} + z_{kj} - z_{ki} \leq 1, \qquad \forall k \in C_{BR}, h, i, j \in V : i \in D_{hj}, \tag{5.23}$$

$$o \in \mathbb{B}^{|V_B|L_B + |V_R|L_R}, \tag{5.24}$$

$$z \in \mathbb{B}^{|V|L_{BR}}. \tag{5.25}$$

Constraints (5.20) ensure that, if i belongs to the convex hull of a group of its opposite class, then i is an outlier (conditions (M1) and (M2)). By (5.19), we have that, for every $i \in V_{BR}$, i must belong to at most one group of its class (condition (M4)). Constraints (5.21) ensure that vertices in V_N can not belong to the intersection of convex sets of opposite classes (condition (M3)). Constraints (5.22) say that each initially classified vertex must be an outlier and/or belong to at least one convex set of its class. The convexity of each set, that is, the convex hull of each group, is ensured by constraints (5.23). The objective function minimizes the number of outliers (condition (M5)).

Actually, ILP2^M is related to ILP1^M as follows.

Proposition 5.2.22 *Let F_1^M and F_2^M be the feasible sets of ILP1^M and ILP2^M, respectively. Then, $F_1^M = \text{proj}_o(F_2^M)$.*

Proof Similar to the proof of Proposition 4.2.15. ■

Propositions 5.2.1 and 5.2.22 imply the correctness of ILP2^M.

Corollary 5.2.23 *Formulation ILP2^M is correct.* ■

We remark that constraints (5.19) and (5.20) can be combined into a smaller group of constraints, as follows.

Proposition 5.2.24 *Let $o \in \mathbb{B}^{|V_B|L_B+|V_R|L_R}$ and $z \in \mathbb{B}^{|V|L_{BR}}$. Then, (o,z) satisfies* (5.19)-(5.20) *if, and only if, (o,z) satisfies*

$$\sum_{k \in C_{K(i)}} o_{ki} - z_{\bar{k}i} \geq L_{K(i)} - 1, \quad \forall i \in V_{BR}, \bar{k} \in C_{\bar{K}(i)} \tag{5.26}$$

Proof First, suppose that (o,z) satisfies all constraints in (5.19)-(5.20). Let $i \in V_{BR}$ and $\bar{k} \in C_{\bar{K}(i)}$. If $z_{\bar{k}i} = 0$, then (5.26) becomes the same as (5.19), and so it is satisfied. If $z_{\bar{k}i} = 1$, by (5.20) we have that $o_{ki} = 1$ for all $k \in C_{K(i)}$. Again, (5.26) is satisfied (at equality).

Now suppose that (o,z) satisfies all constraints in (5.26). Let $i \in V_{BR}$. If $z_{\bar{k}i} = 0$ for all $\bar{k} \in C_{\bar{K}(i)}$, then (5.26) becomes the same as (5.19), and constraints (5.20) are trivially satisfied. If $z_{\bar{k}i} = 1$ for some $\bar{k} \in C_{\bar{K}(i)}$, (5.26) yields that $o_{ki} = 1$ for all $k \in C_{K(i)}$, which implies that (5.19) and (5.20) are satisfied. ■

Besides, the following property ensures that we can aggregate constraints (5.22).

Proposition 5.2.25 *Let $\tilde{F}_2^M$ be the set of points satisfying* (5.19)-(5.21)*,* (5.23)-(5.25) *and*

$$\sum_{k \in C_{K(i)}} z_{ki} + \sum_{k \in C_{K(i)}} o_{ki} \geq L_{K(i)}, \quad \forall i \in V_{BR}. \tag{5.27}$$

Then, $F_2^M \subseteq \tilde{F}_2^M$. Conversely, if $(o,z) \in \tilde{F}_2^M$ then there is $(o',z) \in F_2^M$ such that $\sum_{i \in V_{BR}} \sum_{k \in C_{K(i)}} o_{ki} = \sum_{i \in V_{BR}} \sum_{k \in C_{K(i)}} o'_{ki}$.

Proof Notice that (5.27) is the summation in $k \in C_{K(i)}$ of constraints (5.22). Therefore, $F_2^M \subseteq \tilde{F}_2^M$. Now, let $(o,z) \in \tilde{F}_2^M$. If (o,z) satisfies all constraints in (5.22), we just take o'. So, suppose that there are $i \in V_{BR}$ and $k \in C_{K(i)}$ such that $o_{ki} = z_{ki} = 0$. By (5.19) and (5.27), there is $k' \in C_{K(i)}$ such that $z_{k'i} = o_{k'i} = 1$. Let us obtain o' from o by keeping all components except for $o'_{ki} = 1$ and $o'_{k'i} = 0$. Since $o_{ki} + o_{k'i} = o'_{ki} + o'_{k'i} = 1$, the last equality in the statement follows. For the same reason, (o',z) satisfies (5.19). Besides, $z_{ki} + o'_{ki} = z_{k'i} + o'_{k'i} = 1$ implies that (5.22) holds for (o',z). Since the other linear constraints of ILP2^M do not involve o variables, it remains to show that (o',z) satisfies (5.20), or still that $o'_{ki} \geq z_{\bar{k}i}$ and $o'_{k'i} \geq z_{\bar{k}i}$, for all $\bar{k} \in C_{\bar{K}(i)}$. Since (o,z)

satisfies (5.20), we have $0 = o_{ki} \geq z_{\bar{k}i}$, and so $z_{\bar{k}i} = 0$, for all $\bar{k} \in C_{\bar{K}(i)}$. The desired inequalities then follow. ■

5.2.6 The associated polytope - P_2^M

Let P_2^M be the polytope associated with ILP2^M, that is,

$P_2^M = \text{conv}\{(o,z) \in \mathbb{B}^{|V_B|L_B+|V_R|L_R} \times \mathbb{B}^{|V|L_{BR}} : (5.19)-(5.23)\}.$

In addition, discarding the optional constraints (5.19), we also define the relaxed polytope:

$\bar{P}_2^M = \text{conv}\{o \in \mathbb{B}^{|V_B|L_B+|V_R|L_R} : (5.20)-(5.23)\}.$

Observe that P_2^M and $\bar{P}_2^M$ are equal to P_2 if $L_B = L_R = 1$.

As before, let $e^{ki} \in \mathbb{B}^{|V_B|L_B+|V_R|L_R}$, for each $i \in V_{BR}$ and $k \in C_{K(i)}$, be an unit vector with 1 in the position indexed by k and i. Besides, for ease of notation, let $e^{ki} \in \mathbb{B}^{L_{BR}|V|}$, for each $k \in C_{BR}$ and $i \in V$, also stand for an unit vector with the same feature (but with different dimension).

Proposition 5.2.26 *P_2^M and $\bar{P}_2^M$ are full-dimensional.*

Proof Consider the following $L_B|V_B| + L_R|V_R| + L_{BR}|V| + 1$ vectors in $\mathbb{B}^{L_B|V_B|+L_R|V_R|} \times \mathbb{B}^{L_{BR}|V|}$: $(e,0)$, $(e - e^{kj}, e^{kj})$ for every $j \in V_{BR}$ and $k \in C_{K(j)}$, and (e, e^{kj}) for every $j \in V$ and $k \in C_{BR}$. They belong to P_2^M and are affinely independent, thus showing that P_2^M is full-dimensional. Since $P_2^M \subseteq \bar{P}_2^M$, $\bar{P}_2^M$ is also full-dimensional. ■

Observe that $o_{ki} \geq 0$, for all $i \in V_{BR}$ and $k \in C_{K(i)}$, are dominated by Constraints (5.20). Thus, similarly to Proposition 4.2.20, we obtain the following property.

Proposition 5.2.27 *Let $\pi^T o + \mu^T z \geq \pi_0$ be a facet-defining inequality of P_2^M or $\bar{P}_2^M$. If it is different from $o_{ki} \leq 1$, for all $i \in V_{BR}$ and $k \in C_{K(i)}$, then $\pi \geq 0$. Besides, if $\pi \neq 0$, then $\pi_0 > 0$.*

5.2.7 Relations with P_1^M

By Propositions 2.2.14 and 5.2.22, we have that:

Proposition 5.2.28 $P_1^M = \mathrm{proj}_o(P_2^M)$.

Therefore, according to Proposition 2.2.10, it follows that:

Corollary 5.2.29 *If $\pi^T o \geq \pi_0$ is valid for P_1^M, then it is valid for P_2^M.*

5.2.8 Relations with P_2

We can derive relations between P_2^M with P_2 in the same vein as those obtained between P_1^M and P_1 in Subsection 5.2.3. Here, we recover only those directly involved in the definition of valid inequalities.

Regarding polytope P_2, given a point $(o,z) \in \mathbb{B}^{|V_{BR}|} \times \mathbb{B}^{2|V|}$, we partition $o = (o^B, o^R)$ and $z = (z^B, z^R)$, where ^{B} and ^{R} indicate the components indexed by $i \in V_B$ and $i \in V_R$, respectively. For the ease of the notation, we may write oz, oz^B and oz^R for (o,z), (o^B, z^B) and (o^R, z^R), respectively. Generalizing this notation to P_2^M and a point $(o,z) \in \mathbb{B}^{|V_B|L_B + |V_R|L_R} \times \mathbb{B}^{L_{BR}|V|}$, we use (o^k, z^k), for any $k \in C_{BR}$, to represent the components o_{ki}, for all $i \in V_{BR}$ such that $K(i) = k$, and z_{ki}, for all $i \in V$. When highlighting o^k in the whole vector o, we may write $o = (o^k, o')$, where o' comprises the remaining components. Similarly, we use $z = (z^k, z')$ and $oz = (oz^k, oz')$.

Proposition 5.2.30 *Let $k \in C_B$ and $\bar{k} \in C_R$. Let $\tilde{\mathscr{Q}}^{k\bar{k}} : \mathbb{R}^{|V_{BR}|} \times \mathbb{R}^{2|V|} \to \mathbb{R}^{|V_B|L_B + |V_R|L_R} \times \mathbb{R}^{L_{BR}|V|}$ be the affine transformation such that $\tilde{\mathscr{Q}}^{k\bar{k}}(oz^B, oz^R) = (oz^k, oz^{\bar{k}}, oz')$, where*

$$\begin{bmatrix} oz^k \\ oz^{\bar{k}} \\ o' \\ z' \end{bmatrix} = \begin{bmatrix} I_{|V_B|+|V|} & 0 \\ 0 & I_{|V_R|+|V|} \\ 0 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} oz^B \\ oz^R \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ e \\ 0 \end{bmatrix}. \tag{5.28}$$

Then, $\tilde{\mathscr{Q}}^{k\bar{k}}(P_2) \subseteq P_2^M \subseteq \bar{P}_2^M$.

Proof It suffices to prove that (oz^B, oz^R) feasible for ILP2 implies $(oz^k, oz^{\bar{k}}, oz')$ feasible for ILP2^M. Since $o^{k'} = e$ for all $k' \in C_{BR} \setminus \{k, \bar{k}\}$, inequalities (5.19) are satisfied. This together with $z^{k'} = 0$ for all $k' \in C_{BR} \setminus \{k, \bar{k}\}$ imply that all inequalities (5.20)-(5.21), expect for those related to the pair $(k,\bar{k})$, are trivially satisfied. Besides, each of those inequalities associated with $(k,\bar{k})$ holds, provided that (oz^B, oz^R) satisfies (4.14)-(4.15). Similar arguments can be applied to conclude that constraints (5.22)-(5.23) are trivially satisfied if not related to k or $\bar{k}$, and are due to (4.16)-(4.17), otherwise. ■

Proposition 5.2.31 *If $\sum_{k\in C_{BR}}(\pi^k)^T o^k + \sum_{k\in C_{BR}}(\mu^k)^T z^k \geq \pi_0$ is a valid inequality for P_2^M then $(\pi^k)^T o^B + (\pi^{\bar{k}})^T o^R + (\mu^k)^T z^B + (\mu^{\bar{k}})^T z^R \geq \pi_0 - \sum_{k'\in C_{BR}\setminus\{k,\bar{k}\}}(\pi^{k'})^T e$ is valid for P_2, for every $k \in C_B$ and $\bar{k} \in C_R$.*

Proof Let $k \in C_B$ and $\bar{k} \in C_R$. Let $\tilde{\mathscr{Q}}^{k\bar{k}}$ be defined as in Proposition 5.2.30. Then, $\tilde{\mathscr{Q}}^{k\bar{k}}(P_2) \subseteq P_2^M$, and so $\sum_{k\in C_{BR}}(\pi^k)^T o^k + \sum_{k\in C_{BR}}(\mu^k)^T z^k \geq \pi_0$ is valid for $\tilde{\mathscr{Q}}^{k\bar{k}}(P_2)$. By Proposition 2.2.4, it follows that $(\pi^k)^T o^B + (\pi^{\bar{k}})^T o^R + (\mu^k)^T z^B + (\mu^{\bar{k}})^T z^R \geq \pi_0 - \sum_{k'\in C_{BR}\setminus\{k,\bar{k}\}}(\pi^k)^T e$ is valid for P_2. ■

To identify points in P_2^M or $\bar{P}_2^M$ that are not covered by these affine transformations, it will be convenient to rewrite them by grouping the variables in a different way. Precisely, we can rewrite $\tilde{\mathscr{Q}}^{k\bar{k}}(o^B, o^R, z^B, z^R) = (\mathscr{Q}^{k\bar{k}}(o^B, o^R), \dot{\mathscr{Q}}^{k\bar{k}}(z^B, z^R))$, where $\mathscr{Q}^{k\bar{k}}(o^B, o^R) = (o^k, o^{\bar{k}}, o')$ is given by (5.7) and $\dot{\mathscr{Q}}^{k\bar{k}}(z^B, z^R) = (z^k, z^{\bar{k}}, z')$ is such that

$$\begin{bmatrix} z^k \\ z^{\bar{k}} \\ z' \end{bmatrix} = \begin{bmatrix} I_{|V|} & 0 \\ 0 & I_{|V|} \\ 0 & 0 \end{bmatrix} \begin{bmatrix} z^B \\ z^R \end{bmatrix}. \tag{5.29}$$

Lemma 5.2.32 *Let $i \in V_{BR}$, $k \in C_{K(i)}$ and $\bar{k} \in C_{\bar{K}(i)}$. If $(o,z) \in P_2$ is such that $z_{\bar{K}(i)i} = 0$, then $(\mathscr{Q}^{k\bar{k}}(o) - e^{k'i}, \dot{\mathscr{Q}}^{k\bar{k}}(z) + e^{k'i}) \in \bar{P}_2^M$, for all $k' \in C_{K(i)} \setminus \{k\}$.*

Proof Let $(o,z) \in P_2$ with $z_{\bar{K}(i)i} = 0$. To obtain the claimed result, it suffices to assume that o is integer and show that $(\hat{o},\hat{z}) = (\mathscr{Q}^{k\bar{k}}(o) - e^{k'i}, \dot{\mathscr{Q}}^{k\bar{k}}(z) + e^{k'i})$ satisfies (5.20)-(5.23). Observe that

$$\hat{o}_{cj} = \begin{cases} o_j, & \text{if } c \in \{k,\bar{k}\}, \\ 0, & \text{if } c = k' \text{ and } j = i, \\ 1, & \text{otherwise,} \end{cases} \quad \text{and} \quad \hat{z}_{cj} = \begin{cases} z_{K(i)j}, & \text{if } c = k, \\ z_{\bar{K}(i)j}, & \text{if } c = \bar{k}, \\ 1, & \text{if } c = k' \text{ and } j = i, \\ 0, & \text{otherwise.} \end{cases}$$

By contradiction, first suppose that $(\hat{o},\hat{z})$ violates (5.20) for some $\ell \in V_{BR}$, $c \in C_{K(\ell)}$ and $\bar{c} \in C_{\bar{K}(\ell)}$, i.e. $\hat{o}_{c\ell} = 0$ and $\hat{z}_{\bar{c}\ell} = 1$. Then, we must have $c, \bar{c} \in \{k, \bar{k}, k'\}$, and $\ell = i$ if c or $\bar{c}$ equals k'. On the other hand, since $o_\ell \geq z_{\bar{K}(\ell)\ell}$ by (4.14), we cannot have $\{c, \bar{c}\} = \{k, \bar{k}\}$. Therefore, $\ell = i$, and either $c = k'$ or $\bar{c} = k'$. Since $\bar{c} \in C_{\bar{K}(i)}$ and $k' \in C_{K(i)}$, it must be $c = k'$ and $\bar{c} = \bar{k}$. Then, $1 = \hat{z}_{\bar{c}\ell} = \hat{z}_{\bar{k}\ell} = z_{\bar{K}(i)\ell} = z_{\bar{K}(i)i} = 0$: a contradiction.

Now suppose that $(\hat{o},\hat{z})$ violates (5.21) for some $\ell \in V_N$, $c \in C_B$ and $\bar{c} \in C_R$, i.e. $\hat{z}_{c\ell} = \hat{z}_{\bar{c}\ell} = 1$. Since $\ell \neq i \in V_{BR}$, this means that $\{c, \bar{c}\} = \{k, \bar{k}\}$. Then, by (4.15) $\hat{z}_{c\ell} + \hat{z}_{\bar{c}\ell} = z_{K(i)\ell} + z_{\bar{K}(i)\ell} \leq 1$: a contradiction.

To show that $(\hat{o},\hat{z})$ satisfies (5.22), let $\ell \in V_{BR}$ and $c \in C_{K(\ell)}$. We trivially have $\hat{z}_{c\ell} + \hat{o}_{c\ell} = 1$, if $c \notin \{k, \bar{k}\}$. In the complementary case, $\hat{z}_{c\ell} + \hat{o}_{c\ell} = z_{K(\ell)\ell} + o_\ell \geq 1$, where the inequality follows by (4.16).

Finally, observe that $(\hat{o},\hat{z})$ satisfies constraints (5.23) related to k or $\bar{k}$ due to (4.17). The other constraints in this group are trivially satisfied because z^c, for every $c \notin \{k, \bar{k}\}$, has at most one non-null entry. ■

Lemma 5.2.33 *Let $i \in V_{BR}$, $k \in C_{K(i)}$ and $\bar{k} \in C_{\bar{K}(i)}$. Let $(o,z) \in P_2$ satisfying $o_i = 1$ and $z_{\bar{K}(i)i} = 0$. Then $(\mathscr{Q}^{k\bar{k}}(o) - e^{k'i}, \dot{\mathscr{Q}}^{k\bar{k}}(z)) \in P_2^M$, for all $k' \in C_{K(i)} \setminus \{k\}$.*

Proof Let $(o,z) \in P_2$ with $o_i = 1$ and $z_{\bar{K}(i)i} = 0$. By Lemma 5.2.32, it suffices to assume that (o,z) is integer and show that $(\hat{o},\hat{z}) = (\mathscr{Q}^{k\bar{k}}(o) - e^{k'i}, \dot{\mathscr{Q}}^{k\bar{k}}(z) + e^{k'i})$ satisfies (5.19). Let $\ell \in V_{BR}$. If $\ell \neq i$, then $\sum_{c \in C_{K(\ell)}} \hat{o}_{c\ell} = o_\ell + |C_{K(\ell)}| - 1 \geq L_{K(\ell)} - 1$. If $\ell = i$, then $\sum_{c \in C_{K(\ell)}} \hat{o}_{c\ell} = o_i + |C_{K(\ell)}| - 2 = L_{K(\ell)} - 1$. In both cases, $(\hat{o},\hat{z})$ satisfies (5.19). ■

Lemma 5.2.34 *Let $i \in V$, $k \in C_B$, $\bar{k} \in C_R$ and $k' \in C_{BR} \setminus \{k, \bar{k}\}$. Let $(o,z) \in P_2$. Then, $(\mathscr{Q}^{k\bar{k}}(o), \dot{\mathscr{Q}}^{k\bar{k}}(z) + e^{k'i}) \in P_2^M$ in the following cases:*

1. *$i \in V_{BR}$ and $o_i = 1$; or*
2. *$i \in V_N$, $k' \in C_K$ and $z_{\overline{K}i} = 0$, where $\{K, \overline{K}\} = \{B, R\}$.*

Proof Let (o,z) be an integer point of P_2. We have to show that $(\hat{o},\hat{z}) = (\mathscr{Q}^{k\bar{k}}(o), \dot{\mathscr{Q}}^{k\bar{k}}(z) + e^{k'i})$ satisfies (5.19)-(5.23). Since $\hat{o}^c = e$ for all $c \notin \{k, \bar{k}\}$, constraints (5.19) trivially hold. Every constraint in (5.20)-(5.23) that involves only groups k and $\bar{k}$ also holds because (o,z) satisfies (4.14)-(4.17). Now, let us analyze the cases where at least one involved group is not k or $\bar{k}$. First consider constraint (5.20) related to $\ell \in V_{BR}$, $c \in C_{K(\ell)}$ and $\bar{c} \in C_{\bar{K}(\ell)}$. To get this constraint violated, we should have $\hat{o}_{c\ell} = 0$ and $\hat{z}_{\bar{c}\ell} = 1$. Observe $\hat{o}_{c\ell} = 0$ implies $c \in \{k, \overline{k}\}$, and so $\bar{c} \notin \{k, \overline{k}\}$. Then, $\hat{z}_{\bar{c}\ell} = 1$ leads to $\bar{c} = k'$ and $\ell = i$. Therefore, $i \in V_{BR}$, $k' \in C_{\bar{K}(i)}$, and $0 = \hat{o}_{c\ell} = \hat{o}_{ci} = o_i$. In particular, $i \in V_{BR}$ and $o_i = 0$ show that this constraint is not violated in cases 1-2. Now consider constraint (5.21) for $\ell \in V_N$, $c \in C_K$ and $\bar{c} \in C_{\overline{K}}$, K is such that $k' \in C_K$. Suppose that it is violated, i.e. $\hat{z}_{c\ell} = \hat{z}_{\bar{c}\ell} = 1$. Since $\{c, \bar{c}\} \neq \{k, \overline{k}\}$, it must be $k' = c$, $\ell = i$ and $\bar{c} \in \{k, \overline{k}\}$. Since $i = \ell \in V_N$, the constraint is satisfied in case 1. Besides, $1 = \hat{z}_{\bar{c}\ell} = \hat{z}_{\bar{c}i} = z_{\overline{K}i}$ shows that it is also satisfied in case 2. Finally, constraints (5.22)-(5.23) related to $c \in C_{BR} \setminus \{k, \overline{k}\}$ are satisfied because $\hat{o}^c = e$ and $\hat{z}^c$ has at most one non-null component. ■

We can also show that P_2 is a projection of P_2^M and $\bar{P}_2^M$.

Corollary 5.2.35 *Let $k \in C_B$ and $\overline{k} \in C_R$. Then, $P_2 = \text{proj}_{(oz^k, oz^{\bar{k}})} P_2^M = \text{proj}_{(oz^k, oz^{\bar{k}})} \bar{P}_2^M$.*

Proof Proposition 5.2.30 implies that $P_2 \subseteq \text{proj}_{(oz^k, oz^{\bar{k}})} P_2^M \subseteq \text{proj}_{(oz^k, oz^{\bar{k}})} \bar{P}_2^M$. Now, let $(oz^B, oz^R) \in \text{proj}_{(oz^k, oz^{\bar{k}})} \bar{P}_2^M$. Then, $(oz^k = oz^B, oz^{\overline{k}} = oz^R, oz') \in \bar{P}_2^M$ for some oz'. Since $(oz^k, oz^{\overline{k}}, oz')$ satisfies (5.20)-(5.23), we can deduce that (oz^B, oz^R) satisfies (4.14)-(4.17). Therefore, the integer points in $\text{proj}_{(oz^k, oz^{\bar{k}})} \bar{P}_2^M$ belong to P_2, which implies $\text{proj}_{(oz^k, oz^{\bar{k}})} \bar{P}_2^M \subseteq P_2$. Thus, we get $P_2 = \text{proj}_{(oz^k, oz^{\bar{k}})} P_2^M \subseteq \text{proj}_{(oz^k, oz^{\bar{k}})} \bar{P}_2^M$. ■

5.2.9 Valid inequalities and facets

Corollary 5.2.35 and Proposition 2.2.10 ensure that valid inequalities for P_2 directly yield valid inequalities for $\bar{P}_2^M$ (and consequently for P_2^M).

Proposition 5.2.36 *If $\pi^T oz^B + \lambda^T oz^R \geq \pi_0$ is valid for P_2 then $\pi^T oz^k + \lambda^T oz^{\overline{k}} \geq \pi_0$ is valid for P_2^M and $\bar{P}_2^M$, for all $k \in C_B$ and $\overline{k} \in C_R$.*

In the following, we analyze when the inherited valid inequalities are facet-defining.

Proposition 5.2.37 *If $\pi^T oz^B + \lambda^T oz^R \geq \pi_0$ is facet-defining for P_2 then $\pi^T oz^k + \lambda^T oz^{\overline{k}} \geq \pi_0$ is facet-defining for $\bar{P}_2^M$, for all $k \in C_B$ and $\overline{k} \in C_R$.*

Proof Let $k \in C_B$ and $\overline{k} \in C_R$. Define $C_B' = C_B \setminus \{k\}$ and $C_R' = C_R \setminus \{\bar{k}\}$. Assume that $\pi^T oz^B + \lambda^T oz^R \geq \pi_0$ is facet-defining for P_2. By Proposition 4.2.27, this inequality is different from $o_i \geq 0$ for all $i \in V_{BR}$, $z_{\bar{K}(i)i} \leq 1$ for all $i \in V_{BR}$, and $z_{Ki} \leq 1$ for all $i \in V_N$ and $K \in \{B,R\}$. Let $F = \{(oz^B, oz^R) \in P_2 : \pi^T oz^B + \lambda^T oz^R = \pi_0\}$ and $F^M = \{(oz^k, oz^{\overline{k}}, oz') \in \bar{P}_2^M : \pi^T oz^k + \lambda^T oz^{\overline{k}} = \pi_0\}$. By Proposition 5.2.36, F^M is a face of $\bar{P}_1^M$. Consider the following groups of points (represented in the rows of Table 1):

1. $\tilde{\mathscr{Q}}^{k\overline{k}}(\mathscr{F})$, where $\mathscr{F}$ is a set of $p = |V_{BR}| + 2|V|$ affinely independent points of F. By Proposition 5.2.30, $\tilde{\mathscr{Q}}^{k\overline{k}}(\mathscr{F}) \subseteq P_2^M \subseteq \bar{P}_2^M$. Notice that every point $(oz^k, oz^{\overline{k}}, o', z') = \tilde{\mathscr{Q}}^{k\overline{k}}(oz^B, oz^R) \in \tilde{\mathscr{Q}}^{k\overline{k}}(\mathscr{F})$ has $oz^k = oz^B$, $oz^{\overline{k}} = oz^R$, $o' = e$ and $z' = 0$. They form the rows identified by 1 in the matrix of Table 1.
2. For every $i \in V_{BR}$ and $k' \in C_{K(i)} \setminus \{k, \overline{k}\}$, point $(\mathscr{Q}^{k\overline{k}}(o) - e^{k'i}, \dot{\mathscr{Q}}^{k\overline{k}}(z) + e^{k'i})$, where $oz = (o,z)$ is a solution in F with $z_{\bar{K}(i)i} = 0$ (such a solution exists because $F \neq \{(o,z) \in P_2 : z_{\bar{K}(i)i} = 1\}$). Each of these $|V_B|(L_B - 1) + |V_R|(L_R - 1)$ points belong to $\bar{P}_2^M$ by Lemma 5.2.32. Besides, each of them has the form $(oz^k, oz^{\overline{k}}, o', z')$, where $oz^k = oz^B$, $oz^{\overline{k}} = oz^R$, $o' = e - e^{k'i}$ and $z' = e^{k'i}$. They are represented in the two rows of Table 1 identified by 2.
3. For every $i \in V_{BR}$ and $k' \in C_{BR} \setminus \{k, \overline{k}\}$, point $(\mathscr{Q}^{k\overline{k}}(o), \dot{\mathscr{Q}}^{k\overline{k}}(z) + e^{k'i})$, where $oz = (o,z)$ is a solution in F with $o_i = 1$ (such a solution exists because $F \neq \{(o,z) \in P_2 : o_i = 0\}$). Each of these $|V_{BR}|(L_B + L_R - 2)$ points belong to $P_2^M \subseteq \bar{P}_2^M$ by Lemma 5.2.34(1). Besides, each of them has the form $(oz^k, oz^{\overline{k}}, o', z')$, where $oz^k = oz^B$, $oz^{\overline{k}} = oz^R$, $o' = e$ and $z' = e^{k'i}$. They are represented in the four rows of Table 1 identified by 3.
4. For every $i \in V_N$ and $k' \in C_K \setminus \{k, \overline{k}\}$ $(K \in \{B,R\})$, point $(\mathscr{Q}^{k\overline{k}}(o), \dot{\mathscr{Q}}^{k\overline{k}}(z) + e^{k'i})$, where $oz = (o,z)$ is a solution in F with $z_{\overline{K}i} = 0$ (such a solution exists because $F \neq \{(o,z) \in P_2 : z_{\overline{K}i} = 1\}$). Each of these $|V_N|(L_B + L_R - 2)$ points belong to $P_2^M \subseteq \bar{P}_2^M$ by Lemma 5.2.34(2). Besides, each of them has the form $(oz^k, oz^{\overline{k}}, o', z')$, where $oz^k = oz^B$, $oz^{\overline{k}} = oz^R$, $o' = e$ and $z' = e^{k'i}$. They are represented in the row of Table 1 identified by 4.

Since components $(oz^k, oz^{\overline{k}})$ in all these $|V_B|L_B + |V_R|L_R + |V|L_{BR}$ are always given by a point in F, we conclude that these points belong to F^M. Besides, because of the matrices $E - I$ in group 2 and I in groups 3-4, the points in these groups are affinely independent. The same happens in group 1 since $\mathscr{F}$ is affinely independent. In addition, the (o', z')-components of these later points are always zero (differently from every point in groups 2-4). Therefore, the whole set of points is affinely independent. This shows that F_2^M is a facet of $\bar{P}_2^M$. ■

#	$(oz^k, oz^{\bar{k}})$	$o_{k'i}$		$z_{k'i}$				
		$i \in V_B$	$i \in V_R$	$i \in V_B$		$i \in V_R$		$i \in V_N$
		$k' \in C'_B$	$k' \in C'_R$	$k' \in C'_B$	$k' \in C'_R$	$k' \in C'_B$	$k' \in C'_R$	$k' \in C'_{BR}$
1	$\in F$	E	E	0	0	0	0	0
2	$\in F$	$E-I$	E	I	0	0	0	0
2	$\in F$	E	$E-I$	0	0	0	I	0
3	$\in F$	E	E	I	0	0	0	0
3	$\in F$	E	E	0	I	0	0	0
3	$\in F$	E	E	0	0	I	0	0
3	$\in F$	E	E	0	0	0	I	0
4	$\in F$	E	E	0	0	0	0	I

Table 1 – Points in F_M. The first column indicates the group that the point belongs to. The remaining columns are entries of the point. E is the matrix of ones. '$\in F$' indicates that each row-vector is a point in F.

Corollary 5.2.38 *The following inequalities define facets of $\bar{P}_2^M$.*

1. $o_{ki} \leq 1$, *for every $i \in V_{BR}$ and $k \in C_{K(i)}$;*
2. $z_{ki} \geq 0$, *for every $i \in V_N$ and $k \in C_{BR}$;*
3. $z_{\bar{k}i} \geq 0$, *for every $i \in V_{BR}$ and $\bar{k} \in C_{\bar{K}(i)}$;*
4. $z_{ki} \leq 1$, *for every $i \in V_{BR}$ and $k \in C_{K(i)}$;*
5. $z_{ki} + z_{\bar{k}i} \leq 1$, *for every $i \in V_N$, $k \in C_B$ and $\bar{k} \in C_R$;*
6. $o_{ki} \geq z_{\bar{k}i}$, *for every $i \in V_{BR}$, $k \in C_{K(i)}$, and $\bar{k} \in C_{\bar{K}(i)}$;*
7. $z_{ki} + o_{ki} \geq 1$, *for every $i \in V_{BR}$ and $k \in C_{K(i)}$;*
8. *The corresponding star inequalities* (4.20)*: $\sum_{h \in L} z_{kh} - (|L|-1) z_{ki} \leq 1$, for every $k \in C_{BR}$;*
9. *The corresponding generalized convexity inequalities* (4.22) *related to a complete shortest path $< l_1, q_1, \ldots, l_t, q_t, l_{t+1} >$: $(\sum_i z_{kl_i}) + z_{kj} - (\sum_i z_{kq_i}) \leq 1$, for every $k \in C_{BR}$.*
10. *The corresponding C_4 inequalities* (4.26)*: $\sum_{i \in S_B \cap V_N} (1 - z_{ki}) + \sum_{i \in S_B \cap V_B} o_{ki} + \sum_{i \in S_R \cap V_N} (1 - z_{\bar{k}i}) + \sum_{i \in S_R \cap V_R} o_{\bar{k}i} \geq 2$, for every $k \in C_B$ and $\bar{k} \in C_R$.*

Regarding P_2^M, we can also guarantee facet-heredity from P_2, except for constraints (4.14).

Proposition 5.2.39 *If $\pi^T oz^B + \lambda^T oz^R \geq \pi_0$ is a facet-defining inequality for P_2 different from constraints* (4.14) *for all $i \in V_{BR}$ such that $L_{K(i)} > 1$, then $\pi^T oz^k + \lambda^T oz^{\bar{k}} \geq \pi_0$ is facet-defining* *for P_2^M, for all $k \in C_B$ and $\bar{k} \in C_R$.*

Proof Identical to the proof of Proposition 5.2.37, except that Lemma 5.2.33 instead of Lemma 5.2.32 is used to obtain the points of group 2. Observe that group 2 is only necessary for $i \in V_{BR}$ such that $L_{K(i)} > 1$. To apply Lemma 5.2.33, we claim that, for every such an i, there exists $(o,z) \in F = \{(oz^B, oz^R) \in P_2 : \pi^T oz^B + \lambda^T oz^R = \pi_0\}$ such that $o_i = 1$ and $z_{\bar{K}(i)i}$. Indeed, for every such an i, the inequality is different from (4.14), i.e. $F \neq \{(o,z) \in P_2 : o_i = z_{\bar{K}(i)i}\}$. This implies the existence of an integer point $(o,z) \in F$ such that $o_i > z_{\bar{K}(i)i}$, that is, $o_i = 1$ and $z_{\bar{K}(i)i} = 0$. ■

By applying Proposition 5.2.39 to the facet-defining inequalities of P_2 stated in Subsection 4.2.3.

Corollary 5.2.40 *The following inequalities define facets of P_2^M.*

1. $o_{ki} \leq 1$, *for every $i \in V_{BR}$ and $k \in C_{K(i)}$;*
2. $z_{ki} \geq 0$, *for every $i \in V_N$ and $k \in C_{BR}$;*
3. $z_{\bar{k}i} \geq 0$, *for every $i \in V_{BR}$ and $\bar{k} \in C_{\bar{K}(i)}$;*
4. $z_{ki} \leq 1$, *for every $i \in V_{BR}$ and $k \in C_{K(i)}$;*
5. $z_{ki} + z_{\bar{k}i} \leq 1$, *for every $i \in V_N$, $k \in C_B$ and $\bar{k} \in C_R$;*
6. $z_{ki} + o_{ki} \geq 1$, *for every $i \in V_{BR}$ and $k \in C_{K(i)}$;*
7. *The corresponding induced star inequalities* (4.20)*: $\sum_{h \in L} z_{kh} - (|L|-1) z_{ki} \leq 1$, for every $k \in C_{BR}$;*
8. *The corresponding generalized convexity inequalities* (4.22) *related to a complete shortest path $< l_1, q_1, \ldots, l_t, q_t, l_{t+1} >$: $(\sum_i z_{kl_i}) + z_{kj} - (\sum_i z_{kq_i}) \leq 1$, for every $k \in C_{BR}$.*
9. *The corresponding induced C_4 inequalities* (4.26)*: $\sum_{i \in S_B \cap V_N} (1 - z_{ki}) + \sum_{i \in S_B \cap V_B} o_{ki} + \sum_{i \in S_R \cap V_N} (1 - z_{\bar{k}i}) + \sum_{i \in S_R \cap V_R} o_{\bar{k}i} \geq 2$, for every $k \in C_B$ and $\bar{k} \in C_R$.*

It is worth remarking that, when constraints (5.20) are covered by Proposition 5.2.39 (for $i \in V_{BR}$ such that $L_{K(i)} > 1$), they are not facet-defining for P_2^M. Indeed, in this case, constraints (5.20) are dominated by inequalities (5.26). By their turn, these later inequalities are facet-defining for P_2^M.

Proposition 5.2.41 *For every $i \in V_{BR}$ and $\bar{k} \in C_{\bar{K}(i)}$, inequality* (5.26) *is facet-defining for P_2^M.*

Proof Let $i \in V_{BR}$ with $L_{K(i)>1}$, and $\bar{k} \in C_{\bar{K}(i)}$. Define $F = \{(o,z) \mid \sum_{k \in C_{K(i)}} o_{ki} - z_{\bar{k}i} = L_{K(i)} - 1\}$. Suppose that $F \subseteq F' := \{(o,z) \mid \pi^T o + \lambda^T z = \pi_0\}$. The following items show that F is a facet:

- $\lambda_{kj} = 0$, for $k \neq \bar{k}$ and $j \in V$: it is a consequence of points $(e, e^{\bar{k}i}) \in F$ and $(e, e^{\bar{k}i} + e^{kj}) \in F$.
- $\lambda_{\bar{k}j} = 0$, for $j \neq i$: let $c \in C_{K(i)}$, and so $c \neq \bar{k}$. Then, points $(e - e^{ci}, e^{ci}) \in F$ and $(e - e^{ci}, e^{ci} + e^{\bar{k}j}) \in F$ show that $\lambda_{\bar{k}j} = 0$.
- $\pi_{ki} - \lambda_{\bar{k}i} = 0$, for $k \in C_{K(i)}$: using points $(e - e^{ki}, e^{ki}) \in F$ and $(e, e^{\bar{k}i}) \in F$, we get $\pi_{ki} + \lambda_{ki} + \lambda_{\bar{k}i} = 0$, and so $\pi_{ki} = -\lambda_{\bar{k}i}$, as $k \neq \bar{k}$ and $\lambda_{ki} = 0$ by the first item.
- $\pi_{kj} = 0$, for $j \in V_{K(i)}$ and $k \in C_{K(j)}$: since $k \neq \bar{k}$, $(e - e^{kj}, e^{\bar{k}i} + e^{kj}) \in F$. This point together with $(e, e^{\bar{k}i}) \in F$ and $\lambda_{kj} = 0$ imply $\pi_{kj} = 0$.
- $\pi_{kj} = 0$, for $j \in V_{\bar{K}(i)}$ and $k \in C_{K(j)}$: let $c \in C_{K(i)}$, and so $c \neq k$. The points $(e - e^{ci}, e^{ci}) \in F$ and $(e - e^{ci} + e^{kj}, e^{ci} + e^{kj}) \in F$ imply that $\pi_{kj} = \lambda_{kj}$. So, the first two items lead to $\pi_{kj} = 0$.

■

To complement the study about the constraints of ILP2^M, we show below which of them do not define facets of P_2^M.

Proposition 5.2.42 *The constraints below do not define facets of P_2^M.*

1. $o_{ki} \geq 0$, *for every $i \in V_{BR}$ and $k \in C_{K(i)}$.*
2. $z_{\bar{k}i} \leq 1$, *for every $i \in V_{BR}$ and $\bar{k} \in C_{\bar{K}(i)}$.*
3. $z_{ki} \leq 1$, *for every $i \in V_N$ and $k \in C_{BR}$.*
4. $z_{ki} \geq 0$, *for every $i \in V_{BR}$ and $k \in C_{K(i)}$.*

Proof Similar to the proof of Proposition 4.2.27. ■

5.3 A more compact formulation for the 2-MGC problem

Propositions 5.2.24 and 5.2.25 allow us to rewrite formulation ILP2^M in a more compact way, via the following variable transformation:

$$o_i = \sum_{k \in C_{K(i)}} o_{ki} - L_{K(i)} + 1 \quad \forall i \in V_{BR}. \tag{5.30}$$

Indeed, notice that the objective function (5.18) becomes

$$\sum_{i\in V_{BR}}\sum_{k\in C_{K(i)}} o_{ki} - \sum_{K\in\{B,R\}} |V_K|(L_K-1) = \sum_{K\in\{B,R\}}\sum_{i\in V_K}\Big(\sum_{k\in C_{K(i)}} o_{ki} - (L_K-1)\Big) = \sum_{i\in V_{BR}} o_i.$$

In addition, constraints (5.19), (5.26) and (5.27) respectively turn to be:

$$\begin{aligned} o_i &\in \mathbb{B} \quad \forall i\in V_{BR}\\ o_i &\geq z_{\bar{k}i}, \quad \forall i\in V_{BR}, \bar{k}\in C_{\bar{K}(i)},\\ \sum_{k\in C_{K(i)}} z_{ki} + o_i &\geq 1, \quad \forall i\in V_{BR}. \end{aligned}$$

Therefore, by Propositions 5.2.24 and 5.2.25, ILP2^M can be rewritten as:

$$(\text{ILP3}^M)\ \min \sum_{i\in V_{BR}} o_i \tag{5.31}$$

$$\text{st: } o_i \geq z_{\bar{k}i}, \qquad \forall i\in V_{BR}, \bar{k}\in C_{\bar{K}(i)} \tag{5.32}$$

$$z_{ki} + z_{\bar{k}i} \leq 1, \qquad \forall i\in V_N, k\in C_B, \bar{k}\in C_R, \tag{5.33}$$

$$\sum_{k\in C_{K(i)}} z_{ki} + o_i \geq 1, \qquad \forall i\in V_{BR}, \tag{5.34}$$

$$z_{kh} + z_{kj} - z_{ki} \leq 1, \qquad \forall k\in C_{BR}, h,i,j\in V : i\in D_{hj}, \tag{5.35}$$

$$o \in \mathbb{B}^{|V_{BR}|}, \tag{5.36}$$

$$z \in \mathbb{B}^{|V|L_{BR}}. \tag{5.37}$$

Observe that ILP3^M keeps the same variables z from ILP2^M whereas variables o are used like in ILP1, but with a slightly different meaning. Precisely,

$$o_i = \begin{cases} 1, & \text{if } i\in V_{BR}\setminus \bigcup_{k\in C_{K(i)}} A_k,\\ 0, & \text{if } i\in \bigcup_{k\in C_{K(i)}} A_k, \end{cases}$$

for each $i\in V_{BR}$, and

$$z_{ki} = \begin{cases} 1, & \text{if } i\in H[A_k],\\ 0, & \text{otherwise,} \end{cases}$$

for each $k\in C_{BR}$ and $i\in V$. Note that a feasible solution for ILP3^M may not satisfy condition (M4).

5.3.1 The associated polytope - P_3^M

Let P_3^M be the polytope associated with ILP3^M, that is,

$$P_3^M = \text{conv}\{(o,z)\in \mathbb{B}^{|V_{BR}|}\times\mathbb{B}^{|V|L_{BR}} : (5.32)-(5.35)\}.$$

Remark that $e^i \in \mathbb{B}^{|V_{BR}|}$, for each $i \in V_{BR}$, is a unit vector with 1 in the position indexed by i. Also, vectors $e^{ki} \in \mathbb{B}^{L_{BR}|V|}$, for all $k \in C_{BR}$ and $i \in V$, are the unit vectors with 1 in the position indexed by ki.

Proposition 5.3.1 *P_3^M is full-dimensional.*

Proof Consider the following $|V_{BR}| + L_{BR}|V| + 1$ vectors in $\mathbb{B}^{|V_{BR}|} \times \mathbb{B}^{L_{BR}|V|}$: $(e, 0)$, $(e - e^j, e^{kj})$ for every $j \in V_{BR}$ and an arbitrary $k \in C_{K(j)}$, and (e, e^{kj}) for every $j \in V$ and $k \in C_{BR}$. They belong to P_3^M and are affinely independent, thus showing that P_3^M is full-dimensional. ■

Arguments similar to those used to obtain Proposition 5.2.27 can be employed for P_3^M. Note that $o_i \geq 0$, for all $i \in V_{BR}$, is not facet-defining due to (5.32).

Proposition 5.3.2 *Let $\pi^T o + \mu^T z \geq \pi_0$ be a facet-defining inequality of P_3^M. If it is different from $o_i \leq 1$, for all $i \in V_{BR}$, then $\pi \geq 0$ and $\pi_0 > 0$.*

5.3.2 *Relations with P_2^M*

In this subsection, we derive relations between P_3^M and P_2^M. Toward this end, we define mappings from one polytope to the other one. Some of them are similar to those introduced in Subsection 5.2.3, which relate P_1 and P_1^M.

As in Subsection 5.2.8, given a point $(o, z) \in \mathbb{B}^{|V_B|L_B + |V_R|L_R} \times \mathbb{B}^{L_{BR}|V|}$, we use $oz^k = (o^k, z^k)$, for any $k \in C_{BR}$, to represent the components o_{ki}, for all $i \in V_{BR}$ such that $K(i) = k$, and z_{ki}, for all $i \in V$. To separate components o^k from the remaining ones in o, we may write $o = (o^k, o')$. Similarly, we use $z = (z^k, z')$ and $oz = (oz^k, oz')$. As for a point in P_3^M, given $(o, z) \in \mathbb{B}^{|V_{BR}|} \times \mathbb{B}^{L_{BR}|V|}$, we partition z as before and partition $o = (o^B, o^R)$, where o^B and o^R indicate the components indexed by $i \in V_B$ and $i \in V_R$, respectively.

Unfortunately, we cannot define an affine transformation from P_3^M to P_2^M similar to that one from P_1 to P_1^M (see Proposition 5.2.5) or from P_2 to P_2^M (see Proposition 5.2.30). However, we can still consider the following mapping, which will be useful to relate the polytopes. Let $F_3^M \subset \mathbb{B}^{|V_{BR}|} \times \mathbb{B}^{L_{BR}|V|}$ be the feasible set of ILP3^M.

Proposition 5.3.3 *Let $\mathscr{T} : F_3^M \to \mathbb{R}^{|V_B|L_B+|V_R|L_R} \times \mathbb{R}^{L_{BR}|V|}$ be such that $\mathscr{T}(o,z) = (\tilde{o},z)$, where, for all $i \in V_{BR}$ and $k \in C_{K(i)}$,*

$$\tilde{o}_{ki} = \begin{cases} o_i, & \text{if } k = k_i, \\ 1, & \text{otherwise}, \end{cases} \tag{5.38}$$

and k_i is any fixed element in $C_{K(i)}$ satisfying $z_{k_i i} = 1$ if $o_i = 0$ (such an element exist due to constraints (5.34)*). Then, $\mathscr{T}(F_3^M) \subseteq F_2^M$.*

Proof Let $(o,z) \in F_3^M$ and $(\tilde{o},z) = \mathscr{T}(o,z) \in \mathbb{B}^{|V_B|L_B+|V_R|L_R} \times \mathbb{B}^{L_{BR}|V|}$. Let $i \in V_{BR}$. First, from (5.38), observe that $(\tilde{o},z)$ trivially satisfies (5.19). Regarding the other constraints of ILP2^M, we only have to take care of (5.22) when $k = k_i$, as (o,z) satisfies (5.32)-(5.35) and $\tilde{o}_{ki} = 1$ if $k \neq k_i$. In such a case, we have $z_{ki} + \tilde{o}_{ki} = z_{k_i i} + o_i$, which is at least 1 by hypothesis. Consequently, constraints (5.22) are satisfied. Therefore, $(\tilde{o},z) \in F_2^M$. ■

In the converse direction, we can define an affine transformation similar to that one from P_1^M to P_1 (see Proposition 5.2.10). Moreover, the transformation from P_2^M to P_3^M is now surjective.

Proposition 5.3.4 *Let $\mathscr{T}$ be as defined in Proposition 5.3.3 and $\hat{\mathscr{R}} : \mathbb{R}^{|V_B|L_B+|V_R|L_R} \times \mathbb{R}^{L_{BR}|V|} \to \mathbb{R}^{|V_{BR}|} \times \mathbb{R}^{L_{BR}|V|}$ be the affine transformation such that $\hat{\mathscr{R}}(\tilde{o},z) = (\hat{o},z)$, where*

$$\hat{o}_i = \sum_{k \in C_{K(i)}} \tilde{o}_{ki} - L_{K(i)} + 1 \quad \forall i \in V_{BR}. \tag{5.39}$$

Then $\hat{\mathscr{R}}(\mathscr{T}(o,z)) = (o,z)$, for all $(o,z) \in F_3^M$, and $P_3^M = \hat{\mathscr{R}}(P_2^M)$.

Proof Let $(o,z) \in F_3^M$. By Proposition 5.3.3, $(\tilde{o},z) = \mathscr{T}(o,z) \in F_2^M$. Let $(\hat{o},z) = \hat{\mathscr{R}}(\tilde{o},z)$. By (5.38), for every $i \in V_{BR}$, $\hat{o}_i = \sum_{k \in C_{K(i)}} \tilde{o}_{ki} - L_{K(i)} + 1 = \tilde{o}_{k_i i} = o_i$. Therefore, $\hat{\mathscr{R}}(\mathscr{T}(o,z)) = (o,z)$, and so $F_3^M = \hat{\mathscr{R}}(\mathscr{T}(F_3^M))$. By Proposition 5.3.3, it follows that $F_3^M \subseteq \hat{\mathscr{R}}(F_2^M)$. This implies, by Proposition 2.2.8, that $P_3^M \subseteq \hat{\mathscr{R}}(P_2^M)$.

Now, let $(\hat{o},z) = \hat{\mathscr{R}}(o,z)$, for some integer point $(o,z) \in P_2^M$. We need to show that $(\hat{o},z)$ satisfies (5.32)-(5.35). Since the mapping preserves the z-components, the result is trivial for (5.33) and (5.35). Let $i \in V_{BR}$. If $\hat{o}_i = 1$, constraints (5.32) and (5.34) related to i are clearly satisfied. Now suppose that $\hat{o}_i = 0$. By (5.39), there is $k \in C_{K(i)}$ such that $\tilde{o}_{ki} = 0$. Then, constraints (5.22) and (5.20) imply that $z_{ki} = 1$ and $z_{\bar{k}i} = 0$ for all $\bar{k} \in C_{\bar{K}(i)}$. Therefore, constraints (5.32) and (5.34) are satisfied. ■

5.3.3 Valid inequalities and facets

In general, every valid inequality for P_2^M yields a valid inequality for P_3^M in the same line of Proposition 5.2.12.

Proposition 5.3.5 *If $\sum_{k\in C_{BR}}(\pi^k)^T o^k + \mu^T z \geq \pi_0$ is valid for P_2^M then $\sum_{i\in V_{BR}} \pi_{k_i i} o_i + \mu^T z \geq \pi_0 - \sum_{i\in V_{BR}} \sum_{k\in C_{K(i)}\setminus\{k_i\}} \pi_{ki}$ is valid for P_3^M, where $\pi_{k_i i} = \min\{\pi_{ki} : k \in C_{K(i)}\}$ for all $i \in V_{BR}$.*

Proof Let $(\hat{o},z) \in P_3^M$. By Proposition 5.3.4, there is $(o,z) \in P_2^M$ such that $(\hat{o},z) = \hat{\mathscr{R}}(\tilde{o},z)$, that is, $\hat{o}_i = \sum_{k\in C_{K(i)}} \tilde{o}_{ki} - L_{K(i)} + 1 = \sum_{k\in C_{K(i)}}(\tilde{o}_{ki} - 1) + 1$ for all $i \in V_{BR}$. It follows that

$$
\begin{aligned}
\sum_{i\in V_{BR}} \pi_{k_i i}\hat{o}_i &= \sum_{i\in V_{BR}} \sum_{k\in C_{K(i)}} \pi_{k_i i}(\tilde{o}_{ki} - 1) + \sum_{i\in V_{BR}} \pi_{k_i i} \\
&\geq \sum_{i\in V_{BR}} \sum_{k\in C_{K(i)}} \pi_{ki}(\tilde{o}_{ki} - 1) + \sum_{i\in V_{BR}} \pi_{k_i i} && \text{(due to } \pi_{k_i i} \leq \pi_{ki}, \tilde{o}_{ki} \leq 1) \\
&\geq \pi_0 - \mu^T z - \sum_{i\in V_{BR}} \sum_{k\in C_{K(i)}\setminus\{k_i\}} \pi_{ki} && \text{(due to } \sum_{k\in C_{BR}} (\pi^k)^T \tilde{o}^k + \mu^T z \geq \pi_0).
\end{aligned}
$$

Therefore, $\sum_{i\in V_{BR}} \pi_{k_i i}\hat{o}_i + \mu^T z \geq \pi_0 - \sum_{i\in V_{BR}} \sum_{k\in C_{K(i)}\setminus\{k_i\}} \pi_{ki}$ is valid for P_3^M. ■

For instance, the application of Proposition 5.3.5 to constraints (5.19) produces $o_i \geq 0$. This transformation does not yield strong inequalities, as already observed.

On the other hand, using the rationale of Propositions 5.2.11 and 5.2.31, facet-defining inequalities for P_2^M can be translated into facet-defining inequalities for P_3^M, provided that they are valid inequalities for P_3^M.

Proposition 5.3.6 *Let $\sum_{k\in C_{BR}}(\pi^k)^T o^k + \mu^T z \geq \pi_0$ defines a facet F of P_2^M, $k \in C_B$ and $\bar{k} \in C_R$. Suppose that $\dot{F} = \{(o,z) \in F : o_{k'i} = 1, \forall i \in V_{BR} \forall k' \in K(i)\setminus\{k,\bar{k}\}\} \neq \emptyset$. If $(\pi^k)^T o^B + (\pi^{\bar{k}})^T o^R + \mu^T z \geq \pi_0 - \sum_{k'\in C_{BR}\setminus\{k,\bar{k}\}}(\pi^{k'})^T e$ is valid for P_3^M, then it is facet-defining for P_3^M.*

Proof Since $\dot{F} \neq \emptyset$ and $o_{k'i} \leq 1$ defines a facet of P_2^M (see Corollary 5.2.40), for all $i \in V_{BR}$ and $k' \in K(i)$, we can conclude that $\dim(\dot{F}) = \dim(P_2^M) - |V_B|(L_B - 1) - |V_R|(L_R - 1) - 1$. Therefore, there exists a subset OZ of $\dot{F}$ with $|V_B| + |V_R| + L_{BR}|V|$ affinely independent integer points. Given $(o,z) \in OZ$, let $(\hat{o},z) = \hat{\mathscr{R}}(o,z)$, as defined by Proposition 5.3.4, and $\widehat{OZ} = \hat{\mathscr{R}}(OZ)$. Since

$o_{k'i} = 1$ for all $i \in V_{BR}$ and $k' \in K(i) \setminus \{k, \bar{k}\}$, we have $\hat{o}_i = o_{ki}$ (if $i \in V_B$) and $\hat{o}_i = o_{\bar{k}i}$ (if $i \in V_R$). By constraints (5.22), $z_{ki} + \hat{o}_i = z_{ki} + o_{ki} \geq 1$, for all $i \in V_B$, and $z_{\bar{k}i} + \hat{o}_i = z_{\bar{k}i} + o_{\bar{k}i} \geq 1$, for all $i \in V_R$. Then, we can apply to $\widehat{OZ}$ the mapping $\mathscr{T}$ defined in Proposition 5.3.3 with $k_i = k$, if $i \in V_B$, and $k_i = \bar{k}$, if $i \in V_R$. This way, $\mathscr{T}$ becomes an affine transformation. Recall that $o = (o^k, o^{\bar{k}}, e)$. Then, by (5.38), $\mathscr{T}(\hat{o}, z) = (o, z)$, that is, $\mathscr{T}(\hat{\mathscr{R}}(o, z)) = (o, z)$. As $\mathscr{T}$ is affine in this case and OZ is affinely independent, Proposition 2.2.5 ensures that $\widehat{OZ} = \hat{\mathscr{R}}(OZ)$ is affinely independent. Besides, Proposition 5.3.4 implies that $\widehat{OZ} \subset P_3^M$. We also claim that every point $(\hat{o}, z) \in \widehat{OZ}$ satisfies $(\pi^k)^T \hat{o}^B + (\pi^{\bar{k}})^T \hat{o}^R + \mu^T z = \pi_0 - \sum_{k' \in C_{BR} \setminus \{k, \bar{k}\}} (\pi^{k'})^T e$. Indeed, $(\hat{o}, z) = \hat{\mathscr{R}}(o, z)$ for some $(o, z) \in \dot{F}$. Then, $\hat{o}^B = o^k$, $\hat{o}^R = o^{\bar{k}}$, $o^{k'} = e$ for all $k' \in C_{BR} \setminus \{k, \bar{k}\}$, and $\sum_{k \in C_{BR}} (\pi^k)^T o^k + \mu^T z = \pi_0$. The claimed equality then follows. In addition, since $\hat{o}^B = o^k$, $\hat{o}^R = o^{\bar{k}}$, $o^{k'} = e$ for all $k' \in C_{BR} \setminus \{k, \bar{k}\}$, for every $(o, z) \in OZ$ and $(\hat{o}, z) = \hat{\mathscr{R}}(o, z)$, it follows that $\hat{\mathscr{R}}(o, z) \neq \hat{\mathscr{R}}(o', z)$ whenever $(o, z) \neq (o', z)$. Thus, $|\widehat{OZ}| = |OZ| = |V_B| + |V_R| + L_{BR}|V|$. ■

Corollary 5.3.7 *If $\mu^T z \geq \pi_0$ is facet-defining for P_2^M, then $\mu^T z \geq \pi_0$ is facet-defining for P_3^M.*

Using Proposition 5.3.6, we can show that the following inequalities are facet-defining for P_3^M.

Corollary 5.3.8 *The inequalities below define facets of P_3^M.*

1. $o_i \leq 1$, *for every $i \in V_{BR}$;*
2. $z_{ki} \geq 0$, *for every $i \in V_N$ and $k \in C_{BR}$;*
3. $z_{ki} \geq 0$, *for every $i \in V_{BR}$, $k \in C_{K(i)}$, such that $L_{K(i)} > 1$;*
4. $z_{\bar{k}i} \geq 0$, *for every $i \in V_{BR}$ and $\bar{k} \in C_{\bar{K}(i)}$;*
5. $z_{ki} \leq 1$, *for every $i \in V_{BR}$ and $k \in C_{K(i)}$;*
6. $z_{ki} + z_{\bar{k}i} \leq 1$, *for every $i \in V_N$, $k \in C_B$ and $\bar{k} \in C_R$;*
7. $o_i \geq z_{\bar{k}i}$, *for every $i \in V_{BR}$ and $\bar{k} \in C_{\bar{K}(i)}$;*
8. *The corresponding induced star inequality* (4.20)*: $\sum_{h \in L} z_{kh} - (|L| - 1) z_{ki} \leq 1$, $k \in C_{BR}$;*
9. *The corresponding generalized convexity inequality* (4.22) *related to a complete shortest path $< l_1, q_1, \ldots, l_t, q_t, l_{t+1} >$: $(\sum_i z_{kl_i}) + z_{kj} - (\sum_i z_{kq_i}) \leq 1$, $k \in C_{BR}$.*
10. *The corresponding induced C_4 inequality* (4.26)*: $\sum_{i \in S_B \cap V_N} (1 - z_{ki}) + \sum_{i \in S_B \cap V_B} o_i + \sum_{i \in S_R \cap V_N} (1 - z_{\bar{k}i}) + \sum_{i \in S_R \cap V_R} o_i \geq 2$, for every $k \in B$ and $\bar{k} \in R$.*

We remark that the validity hypothesis used in Proposition 5.3.6 does not always hold. For instance, consider constraints (5.22). Its direct translation to P_3^M would be $z_{ki} + o_i \geq 1$, which is not valid. Actually, the corresponding constraints in ILP3^M are (5.34), which we show to be facet-defining next.

Proposition 5.3.9 *For every $i \in V_{BR}$, constraint* (5.34) *is facet-defining for P_3^M.*

Proof Let $i \in V_{BR}$. Consider the following points:

$$\begin{cases} (e,0), & \\ (e-e^i, e^{ki}), & \text{for each } k \in C_{K(i)}, \\ (e-e^j, e^{cj}), & \text{for each } j \in V_{BR} \setminus \{i\}, \text{ and for an arbitrary } c \in C_{K(j)}, \\ (e, e^{kj}), & \text{for each } j \in V \setminus \{i\}, k \in C_{K(i)}, \\ (e, e^{\bar{k}j}), & \text{for each } j \in V, \bar{k} \in C_{\bar{K}(i)}. \end{cases}$$

These are $|V_{BR}| + |V|L_{BR}$ affinely independent points satisfying $\sum_{k \in C_{K(i)}} z_{ki} + o_i \geq 1$ at equality. Therefore, this inequality is facet-defining. ■

To complement the analyzes of the constraints of ILP3^M, we present the following remark.

Proposition 5.3.10 *The constraints below do not define facets of P_3^M.*

1. $o_i \geq 0$*,* $i \in V_{BR}$*,*
2. $z_{\bar{k}i} \leq 1$*,* $i \in V_{BR}$ *and* $\bar{k} \in \bar{K}(i)$*,*
3. $z_{ki} \leq 1$*,* $i \in V_N$ *and* $k \in C_{BR}$*.*

Proof Note that $o_i = 0$ implies $z_{\bar{k}i} = 0, \forall \bar{k} \in \bar{K}(i)$, due to (5.32). So, $o_i \geq 0$ and $z_{\bar{k}i} \leq 1$ can not be facet-defining. Finally, observe that $z_{ki} \leq 1$ is dominated by constraint $z_{ki} + z_{li} \leq 1$, where l is an arbitrary group of the opposite class. ■

We now analyze the transformation of valid inequalities in the opposite direction, that is, from P_3^M to P_2^M.

Proposition 5.3.11 *If $\pi^T o^B + \lambda^T o^R + \mu^T z \geq \pi_0$ is valid for P_3^M, then $\sum_{k \in C_B} \pi^T o^k + \sum_{\bar{k} \in C_R} \lambda^T o^{\bar{k}} + \mu^T z \geq \pi_0 + (L_B - 1)\pi^T e + (L_R - 1)\lambda^T e$ is valid for P_2^M.*

Proof Let $\pi^T o^B + \lambda^T o^R + \mu^T z \geq \pi_0$ be valid for P_3^M. Since $P_3^M = \hat{\mathscr{R}}(P_2^M)$ according to Proposition 5.3.4, it follows from Proposition 2.2.3 and (5.39) that

$$\sum_{i \in V_B} \pi_i \left(\sum_{k \in C_B} o_{ki} - L_B + 1 \right) + \sum_{i \in V_R} \lambda_i \left(\sum_{\bar{k} \in C_R} o_{ki} - L_R + 1 \right) + \mu^T z \geq \pi_0 \tag{5.40}$$

is valid for P_2^M. This is exactly the desired inequality. ■

It is worth comparing Proposition 5.3.11 with Propositions 5.2.13 and 5.2.36. Let us consider the inequality $\pi^T o^k + \lambda^T o^{\bar{k}} + \mu^T z \geq \pi_0$, for $k \in C_B$ and $\bar{k} \in C_R$, which is directly obtained from a valid inequality $\pi^T o^B + \lambda^T o^R + \mu^T z \geq \pi_0$ for P_3^M. Now, this resulting inequality is dominated by (5.40), whenever $\pi \geq 0$ and $\lambda \geq 0$ (by Proposition 5.3.2, these conditions hold if $\pi^T o^B + \lambda^T o^R + \mu^T z \geq \pi_0$ differs from $o_i \leq 1$ and defines a facet of P_3^M). Indeed, it is the summation of (5.40) with the valid inequalities $\pi_i o_{k'i} \leq (L_B - 1)\pi_i$, for all $i \in V_B$ and $k' \in C_B \setminus \{k\}$, and $\lambda_i o_{k'i} \leq (L_R - 1)\lambda_i$, for all $i \in V_R$ and $k' \in C_R \setminus \{\bar{k}\}$. Note that these two later groups comprise valid inequalities because $o_{k'i} \leq 1$, $\pi_i \geq 0$, $\lambda_i \geq 0$, and they are used only if $k' \in C_{K(i)} \setminus \{k, \bar{k}\}$, which implies $L_{K(i)} \geq 2$.

As an example, consider a constraint (5.32) from ILP3^M as $\pi^T o^B + \lambda^T o^R + \mu^T z \geq \pi_0$. For $k \in C_B$ and $\bar{k} \in C_R$, the corresponding inequality $\pi^T o^k + \lambda^T o^{\bar{k}} + \mu^T z \geq \pi_0$ is a constraint (5.20) of ILP2^M, which is dominated by (5.26). This later inequality is exactly (5.40) in this case.

5.3.4 Summary of facetness results

For a better overview of the facetness results, we summarize them in the following tables and Figure 23:

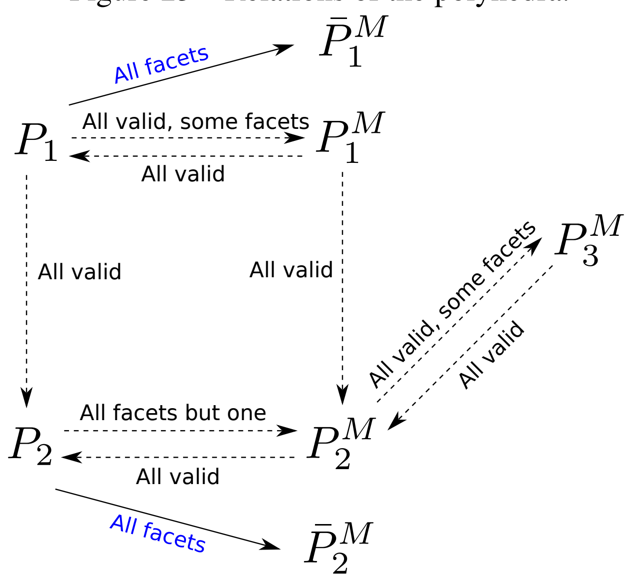
Constraint of P_1^M	**Facet-defining?**
$o_{ki} \leq 1$	yes
$o_{ki} \geq 0$	yes, iff $\lvert C_{K(i)} \rvert = 1$
$\sum_{k \in C_{K(i)}} o_{ki} \geq L_{K(i)} - 1$	yes
$\sum_{j \in S} o_{\bar{k}j} + \sum_{k \in C_{K(i)}} o_{ki} \geq L_{K(i)}$	yes, if (F1) and (F2) hold
Generalized C_4 inequalities	yes, if $L_B = L_R = 1$
Generalized 3-path inequalities	yes, if (F1) and (F2) hold
X-swing inequalities	yes, if (F1) and (F2) hold
Star tree inequalities	yes, if (F1) and (F2) hold
Alternating path inequalities	yes, if (F1) and (F2) hold

Constraint of $\bar{P}_1^M$	**Facet-defining?**
$o_{ki} \leq 1$	yes
$o_{ki} \geq 0$	yes, iff $\lvert C_{K(i)} \rvert = 1$
Generalized C_4 inequalities	yes
Generalized 3-path inequalities	yes, if (F1) and (F2) hold
X-swing inequalities	yes, if (F1) and (F2) hold
Star tree inequalities	yes, if (F1) and (F2) hold
Alternating path inequalities	yes, if (F1) and (F2) hold

Constraint of P_2^M	**Facet-defining?**
$o_{ki} \leq 1$	yes
$z_{ki} \geq 0, i \in V_N$	yes
$z_{\bar{k}i} \geq 0, i \in V_{BR}$	yes
$z_{ki} \leq 1, k \in C_{K(i)}$	yes
$z_{ki} + z_{\bar{k}i} \leq 1$	yes
$z_{ki} + o_{ki} \geq 1$	yes
$o_{ki} \geq z_{\bar{k}i}$	yes, iff $\lvert C_{K(i)} \rvert = 1$
$\sum_{k \in C_{K(i)}} o_{ki} - z_{\bar{k}i} \geq L_{K(i)} - 1$	yes
Star inequalities	yes, if the star is induced
Generalized convexity inequalities	yes, if it induces a complete path of G
C_4 inequalities	yes, if the C_4 is induced
$\sum_{k \in C_{K(i)}} o_{ki} \geq L_{K(i)} - 1$	no
$o_{ki} \geq 0$	no
$z_{ki} \geq 0, k \in C_{K(i)}$	no
$z_{\bar{k}i} \leq 1$	no
$z_{ki} \leq 1, i \in V_N$	no

Constraint of P_3^M	Facet-defining?
$o_i \leq 1$	yes
$z_{\bar{k}i} \geq 0, i \in V_{BR}$	yes
$z_{ki} \geq 0, i \in V_N$	yes
$z_{ki} \leq 1, k \in C_{K(i)}$	yes
$z_{ki} \geq 0, k \in C_{K(i)}$	yes, iff $\lvert C_{K(i)} \rvert > 1$
$z_{ki} + z_{\bar{k}i} \leq 1$	yes
$o_i \geq z_{\bar{k}i}$	yes
$\sum_{k \in C_{K(i)}} z_{ki} + o_i \geq 1$	yes
Star inequalities	yes, if the star is induced
Generalized convexity inequalities	yes, if it induces a complete path of G
C_4 inequalities	yes, if the C_4 is induced
$o_i \geq 0$	no
$z_{\bar{k}i} \leq 1, i \in V_{BR}$	no
$z_{ki} \leq 1, i \in V_N$	no

Figure 23 – Relations of the polyhedra.

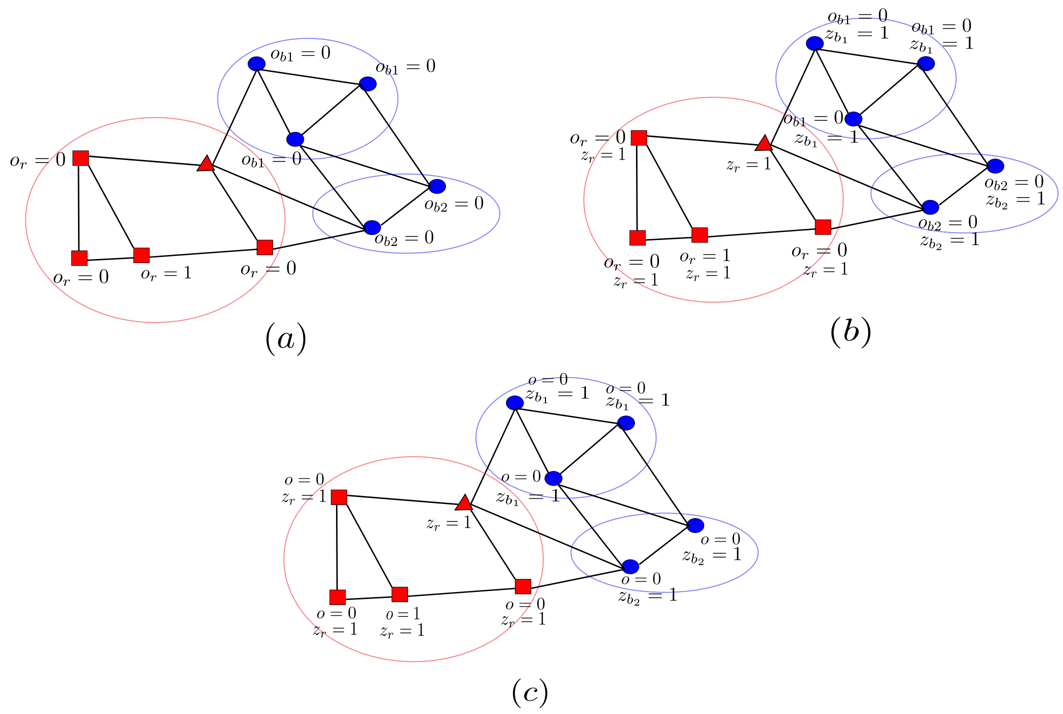


5.3.5 *Comparison of the formulations*

Formulations ILP1^M, ILP2^M and ILP3^M are different from each other, but they are equivalent in the sense that they are all correct for the 2-MGC problem: for each feasible solution of one formulation, there is a corresponding feasible solution in each other formulations such that the outlier vertices, the blue groups and the red groups are the same. Thus, corresponding solutions have equal objective function values and equally classify the vertices in $V_N \setminus V_A$. V_A is the set of all vertices $u \in V_N$ such that u does not belong to the convex hull of any group.

Note that this notion of equivalence says nothing about vertices in V_A. It is because any class can be assigned to them without violating a constraint of the classification problem. An illustration of equivalent solutions for ILP1^M, ILP2^M and ILP3^M is shown in Figure 24. Examples (a), (b) and (c) represent a solution of ILP1^M, ILP2^M and ILP3^M, respectively, for a graph instance with $C_B = \{b_1, b_2\}$ and $C_R = \{r\}$. Circles are vertices of V_B, squares are vertices of V_R, and triangles are vertices of V_N. Close to each vertex, we present the values of the o-variables and z-variables that are relevant to characterize the solution configuration. Precisely, we show only non-null z-variables and, if a vertex has more than one associated o-variable, we show the only that is null.

Figure 24 – An illustration of corresponding solutions.



The set of variables in ILP1^M is included in that of ILP2^M. So, the latter is an extended ILP formulation ((VANDERBECK; WOLSEY, 2010)). It is known that extended formulations may offer some computational advantages. ILP3^M can be seen as an intermediate formulation (between ILP1^M and ILP2^M), since it aggregates the o-variables related to the same vertex and uses the same z-variables included in ILP2^M.

Although the number of constraints in ILP1^M is potentially exponential (this number is polynomial in the other two formulations), the convexity constraints (5.3)-(5.4) can be separated of integer solution in polynomial time (details will be presented in Subsection 6.4.4). However, separation of a fractional solution is NP-hard as it relates to the Hull Number problem ((DOURADO *et al.*, 2009b)). For the single-group case, we could verify by computational experiments tests with small instances that the lower bound provided by the linear relaxation of ILP1 is usually better.

In general, to compare the lower bounds provided by the linear relaxations of the three formulations, first recall that ILP3^M was obtained from ILP2^M by the variable transformation $o_i = \sum_{k \in C_K(i)} o_{ki} - L_{K(i)} + 1$, for each $i \in V_{BR}$. So the linear relaxation lower bound of ILP2^M and ILP3^M are the same. Regarding ILP1^M, by computational experiments, we found instances where the linear relaxation lower bound of ILP1^M was better than the linear relaxation lower bound of ILP2^M as well as instances where opposite relation holds. Thus, the linear relaxation of ILP1^M does not dominate and it is not dominated by the other two formulations. Moreover, it seems that there is no simple theoretical relation between their lower bounds.

However, for the cases with $L_B > 1$ and $L_R > 1$, we can realize that the linear relaxation lower bound is equal to zero for all three formulations. This is because we can just obtain a solution in the following manner. For each $i \in V_{BR}$, set $o_{ki} = o_{k'i} = 0.5$, for arbitrary $k, k' \in C_{K(i)}$, and $o_{li} = 1$ for the remaining $l \in C_{K(i)} \setminus \{k, k'\}$, which leads to $o_i = 0$ in the case of ILP3^M. Besides, set $z_{ki} = z_{k'i} = 0.5$ and null any other z-variables. Thus, we obtain a relaxed feasible solution with value equal to 0. Thus, the larger is the number of allowed groups, the weaker is the linear relaxation, which suggests that more effort has to be spent to develop an efficient branch-and-cut.

Another weakness in the multi-group case is the presence of symmetries due to group indices, which could be removed, leading to strengthened formulations at the expertise of adding some symmetry-breaking constraints. In ILP1^M and ILP2^M, we can also have more than one value assignment for the o-variables that define the same convex sets (and thus the same classification while keeping the number of outliers). In ILP3^M, there are multiple group symmetries because there is no limitation for the number of groups that an initially classified vertex can belong to, since there is no constraint like $\sum_{k \in C_K(i)} o_{ki} \geq L_{K(i)} - 1$ of ILP1^M.

6 ALGORITHMS AND COMPUTATIONAL EXPERIMENTS

In this chapter, we describe the algorithms and report our computational experiments. We aim at analyzing two main aspects: the efficiency of the formulations and the efficiency of the derived facet-defining inequalities as well as the accuracy of the solution of the classification problem.

We focus on the single-group case. We present the algorithms that we developed to solve ILP1 and ILP2. Observe that we can think of several different strategies to include the constraints and valid inequalities in the models. Concerning accuracy, we compare our approach with two known classic algorithms for the Euclidean version of the classification problem.

To test the algorithms, three sets of instances were used in the experiments, namely random instances, realistic instances and synthetic instances. The computational experiments were performed on a machine with an Intel i7-7700 3.6 GHz 8 cores processor, 32 GB RAM and 64 bits Linux OS. The implementation was made using programming language *C++* and version 12.8 of the `CPLEX` package (with default parameters) to solve the linear and integer programming models.

6.1 Calculation of D_{hj}

The set D_{hj} comprises all internal vertices in all shortest paths from h to j. To calculate all D_{hj} sets for a graph instance $G = (V, E)$, we applied a breadth-first search algorithm for each source $h \in V$. In such an algorithm, each vertex j has its D_{hj} set updated every time an adjacent vertex i of j (i.e., $(i, j) \in E$) with lower distance to h ($\delta(h,i) < \delta(h, j)$) is reached. When it happens, the set D_{hj} is updated in the following way: $D_{hj} \leftarrow D_{hj} \cup \{i\} \cup D_{hi}$. So, for a given source $h \in V$, the set D_{hj}, for all $j \in V$, are determined in time $O(n+m)$. Therefore, the complexity to calculate D_{hj} for all $h \in V$ is $O(n(n+m))$. Algorithm 1 shows the corresponding pseudocode.

6.2 Convex hull calculation

To calculate a convex hull W of a subset S, we start with $W' = S$ and iteratively update it. At each iteration, we add to W' all vertices in D_{uv}, for every pair $u, v \in W'$ (with at least one of them added to W' in the previous iteration). This step is repeated until W' does not change. At this point, we get $W = W'$. Sets D_{uv} can be determined a priori with the BFS-like

Algorithm 1: D_{hj} calculating algorithm

Data: Graph Instance G.
Result: All D_{hj} sets.

$D \leftarrow \emptyset$
foreach $h \in V$ **do**
    foreach $j \in V$ **do**
        $d[h][j] \leftarrow \infty$
    $d[h][h] \leftarrow 0$
    Queue $\leftarrow \{h\}$
    repeat
        $i \leftarrow$ Queue.front()
        Queue.popFront()
        foreach $j \in V$ **do**
            if $j \neq i$ *and* $(i, j) \in E$ *and* $d[h][i] < d[h][j]$ **then**
                if $d[h][j] = \infty$ **then**
                    Queue.pushBack(j)
                $d[h][j] \leftarrow d[h][i] + 1$
                if $d[h][j] > 1$ **then**
                    $D_{hj} \leftarrow D_{hj} \cup \{i\} \cup D_{hi}$
    until $Queue = \emptyset$
return D

algorithm described in Algorithm 1.

6.3 Pre-processing

For a given graph instance G, if a vertex $i \in V_N$ neither belongs to the convex hull of V_B nor to the convex hull of V_R, then i can never be reached by any blue or red group in a feasible solution. In this case, we eliminate i from G (i.e., remove i and all incident edges to i), since it can be set to any class.

Now, if a vertex $i \in V_N$ does not belong to the convex hull of V_R but it belongs to the convex hull of V_B, then we know that i can never be reached by any red group in a feasible solution. However, it can still be reached by a blue group, so we fix all z-variables associated with i and red groups: $z_{ri} = 0$, $r \in C_R$, for ILP2. A similar fixing can be done in the case in which i belongs to the convex hull of V_R but does not belong to the convex hull of V_B. To do so, we only need to calculate the convex hulls $H[V_B]$ and $H[V_R]$ and check these conditions.

6.4 Separation algorithms

Although our computational experiments focus on the single-group case, the separation algorithms presented in this section are generalized for the multi-group version of the geodesic classification problem.

6.4.1 C_4 inequalities separator

A separation algorithm for inequalities (4.10) and (4.25) can be obtained by just enumerating and storing in a list and then checking for violation. This list can be created during the construction of the initial integer linear programming model by enumerating all 2-sized subsets of V_{BN} and V_{RN} and verifying the subset pairs that satisfy the requirements of the corresponding inequality. For the sake of time efficiency, our implementation seeks for pairs $(\{i,i'\},\{j,j'\})$ such that $j,j' \in D_{ii'}$ and $i,i' \in D_{jj'}$, instead of pairs such that $j,j' \in H[\{i,i'\}]$ and $i,i' \in H[\{j,j'\}]$. This way we are possibly not considering all inequalities (4.10) or (4.25).

In practice, searching in such a list, instead of enumerating all pairs each time the separator runs, drastically decreases the algorithm's overall running time, since the list size is, on average, much smaller than the worst-case (it is around 0.01% of n^4).

6.4.2 Generalized convexity inequalities separator

We developed a clever separation algorithm for the generalized convexity inequalities (4.21), including the case $t = 1$. The idea is the following: given $h \in V$ and $k \in C_{BR}$, find a complete or incomplete shortest path from h in G whose corresponding inequality yields the maximum value in the left-hand side of (4.21). We search for such a path in a dag (direct acyclic graph) composed by all complete and incomplete shortest paths starting at root $h \in V$ to any other vertex in G. To create such a dag for a given vertex $h \in V$, we apply the steps below:

1. Create the dag of all complete shortest paths starting at h. It is easily calculated by a breadth-first search algorithm similar to Algorithm 1 (Figure 26);
2. Complement the dag of Step 1 by including all necessary arcs to form all incomplete shortest paths starting at h, that is, calculating the transitive hull (i.e., sequences including all possible generalized convexity inequalities with source h). To do so, we apply the known Dijkstra's algorithm (Figure 27);
3. Finally, create a super sink f and add arcs (v,f), for all $v \in V(Dag[h])$ such that $\delta(h,v) > 1$

(Figure 28).

The resulting dag, for a given vertex h, is denoted by $Dag[h]$.

Figure 25 – Graph instance example.

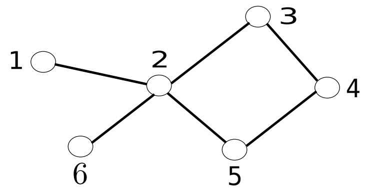


Figure 26 – Separation algorithm for the example of Figure 25. Step 1.

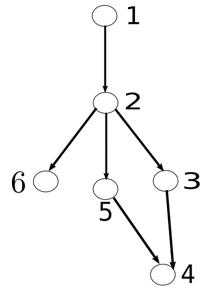


Figure 27 – Separation algorithm for the example of Figure 25. Step 2.

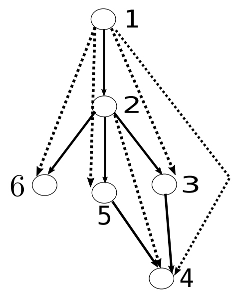
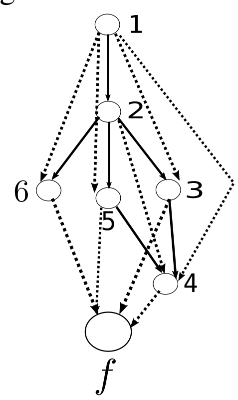


Figure 28 – Separation algorithm for the example of Figure 25. Step 3.



Let us consider an even path in $Dag[h]$, from h to f, whose sequence of vertices is $S = < h = v_1, v_2, \ldots, v_{2p-1}, v_{2p} = f >$. The left-hand side of the generalized convexity inequality (4.21) induced by S and group $k \in C_{BR}$ is

$$z_{kv_1} - z_{kv_2} + z_{kv_3} \dots - z_{kv_{2p-2}} + z_{kv_{2p-1}} = z_{kh} + \frac{z_{kv_1} - z_{kv_2}}{2} - \frac{z_{kv_2} + z_{kv_3}}{2} + \dots - \frac{z_{kv_{2p-2}} + z_{kv_{2p-1}}}{2}. \tag{6.1}$$

Thus, to calculate the most violated path in $Dag[h]$, related to group $k \in C_{BR}$, we assign to each arc (u,v) of $Dag[h]$ a weight $w(u,v) = \frac{z_{ku} - z_{kv}}{2}$. When calculating the weight of a path containing an arc (u,v), we multiply $w(u,v)$ by 1 or -1 depending on u appearing in an odd or even position in the path, respectively. The weight of the path is this signed sum plus $\frac{z_{kh}}{2}$.

We can then determine the maximum weighted path by traversing $Dag[h]$ in the topological order and calculating the paths weights. We just have to take care to multiply the arc weights by 1 or -1 accordingly. In the end, if the weight of the maximum weighted path is greater than 1, the inequality induced by k and this path is violated; otherwise, no violation exists for k.

Algorithm 2 describes the generalized convexity inequalities separator just explained. For each $h \in V$, after computing $Dag[h]$ and adding the artificial sink vertex f, we run a topological sorting algorithm for $Dag[h]$. The array pathWeightPos (resp., pathWeightNeg) saves the maximum weight path from h to each vertex $v \in V$ in $Dag[h]$ among the paths where v appears in an odd (resp. even) position, i.e. the coefficient of z_{kv} is multiplied by 1 (resp. -1) to set the path weight. The parentPos and parentNeg arrays save the predecessors of each vertex in the maximum weight path determined in pathWeightPos and pathWeightNeg, respectively. So, lines 14 - 28 calculate a maximum weight path in $Dag[h]$ from h to the artificial super sink f that yields to a valid generalized convexity inequality using the topological order. Finally, if the corresponding inequality of the path is violated, then it is saved to be added to the integer linear programming model (lines 29 - 31).

The complexity of such a separation algorithm is mainly determined by the Dijkstra's algorithm, which has complexity $O(n^2 \log n))$. The topological sorting algorithm can be implemented within complexity $O(n+m)$ with a depth-first search algorithm. Lines 6 - 31 has three *for* commands which runs in $O(|C_{BR}| \cdot n^2)$ time. Thus, the overall complexity of Algorithm 2 is $O(n^2 \log n + (|C_{BR}| \cdot n^2)) = O(n^2 \log n)$.

Algorithm 2: Generalized convexity inequality separation algorithm

Data: Graph instance G and continuous solution z.

Result: A set of violated generalized convexity inequalities, one for each pair (h,k), $h \in V$, $k \in C_{BR}$.

```
F ← ∅
foreach h ∈ V do
    Compute Dag[h] using depth-first search and Dijkstra's algorithms
    Create a sink node f and add arcs (i, f) for each i ∈ V\{h}
    Run the topological order algorithm for Dag[h]
    foreach k ∈ C_BR do
        foreach u ∈ V(Dag[h]) do
            parentPos[u] ← −1
            parentNeg[u] ← −1
            pathWeightPos[u] ← −∞
            pathWeightNeg[u] ← −∞
        parentPos[h] ← h
        pathWeightPos[h] ← 0
        foreach v ∈ V(Dag[h]) in the topological order do
            if pathWeightPos[v] = −∞ and pathWeightNeg[v] = −∞ then
                continue
            foreach w ∈ V(Dag[h])\{f} do
                if w ≠ v and (v, w) ∈ A(Dag[h]) then
                    if parentPos[v] ≠ −1 and (pathWeightPos[v] + ((z_kv − z_kw)/2) >
                     pathWeightNeg[w]) then
                        pathWeightNeg[w] ← pathWeightPos[v] + ((z_kv − z_kw)/2)
                        parentNeg[w] ← v
                    if parentNeg[v] ≠ −1 and (pathWeightNeg[v] + ((−z_kv + z_kw)/2) >
                     pathWeightPos[w]) then
                        pathWeightPos[w] ← pathWeightNeg[v] + ((−z_kv + z_kw)/2)
                        parentPos[w] ← v
            if v ≠ f and (v, f) ∈ A(Dag[h]) and parentPos[v] ≠ −1 and
             (pathWeightPos[v] + ((z_kv + z_kh)/2) > pathWeightNeg[f]) then
                pathWeightNeg[f] ← pathWeightPos[v] + ((z_kv + z_kh)/2)
                parentNeg[f] ← v
        maxPathWeight ← pathWeightNeg[f]
        if maxPathWeight > 1 then
            determine the violated inequality using parentPos and parentNeg arrays
            add the violated inequality to F
return F
```

6.4.3 Generalized alternating path inequalities separator

The generalized alternating path inequalities (4.12) are the counterparts, for ILP1, of the generalized convexity inequalities. Indeed, for a sequence $< h = l_1, q_1, \ldots, q_t,$ $l_{t+1} = j >$, $k \in C_{K(h)}$ and $\bar{k} \in C_{\bar{K}(h)}$, inequality (4.12) for the multi-group case is $o_{kl_1} + o_{\bar{k}q_1} + o_{kl_2} + o_{\bar{k}q_2} + \ldots + o_{kl_t} + o_{\bar{k}q_t} + o_{kl_{t+1}} \geq t$. It can be rewritten as

$$(1 - o_{kl_1}) - o_{\bar{k}q_1} + (1 - o_{kl_2}) - o_{\bar{k}q_2} + \ldots + (1 - o_{kl_t}) - o_{\bar{k}q_t} + (1 - o_{kl_{t+1}}) \leq 1. \tag{6.2}$$

Also, recall that two consecutive vertices in the sequence have opposite classes.

Using expression (6.2), the separation of inequalities (4.12) can be obtained by a procedure similar to that one presented in Subsection 6.4.2. The two differences are:

- After step 2 (and before step 3), keep in $Dag[h]$ only the arcs that link two initially classified vertices of opposite classes. Remove the other arcs and the vertices that become isolated.
- The weight of a remaining arc (u, v) is $w(u, v) = \frac{(1-o_{ku})-o_{\bar{k}v}}{2}$, if $K(u) = K(v)$, and $w(u, v) = \frac{o_{\bar{k}u}+(1-o_{kv})}{2}$, otherwise. It is important to observe that now we just need to compute a maximum weighted path from h to f (without multiplying the arc weights by 1 or -1).

In practice, we observed that separating all such inequalities was not rewarding. Instead, we restricted ourselves to separating the generalized 3-path inequalities (the special case of (4.12) when $t = 2$) by enumeration.

6.4.4 Lazy constraints separator for elementary $\mathscr{N}$-set inequalities

Since there can be many constraints in ILP1^M, we designed a lazy constraints algorithm to separate integer solutions of ILP1^M. We just search for subsets $S \subseteq V_B$ and $T \subseteq V_R$ whose corresponding $\mathscr{N}$-set elementary constraint is violated. To do so, it is sufficient to transform the current integer solution o of ILP1^M into the corresponding integer solution $(\hat{o}, \hat{z})$ of ILP2^M, by explicitly calculating the convex hull of all groups $A_k = \{j \in V_K \mid o_{kj} = 0\}$, for all $k \in C_K$, $K \in \{\mathbb{B}, R\}$. Then, we verify:

- If $\sum_{k \in C_{K(i)}} \hat{o}_{ki} - L_{K(i)} - 1 < \hat{z}_{\bar{k}i}$, for some $i \in V_{BR}$ and $\bar{k} \in C_{\bar{K}(i)}$, then the constraint $\sum_{j \in S} o_{\bar{k}j} + \sum_{k \in C_{K(i)}} o_{ki} \geq L_{K(i)}$, for $S = A_{\bar{k}}$, is violated (note that this condition implies $\sum_{k \in C_{K(i)}} o_{ki} = L_{K(i)} - 1$ and $\sum_{j \in A_{\bar{k}}} o_{\bar{k}j} = 0$, with $A_{\bar{k}} \neq \emptyset$).

- If $\hat{z}_{ki} + \hat{z}_{\bar{k}i} > 1$, for some $i \in V_N, k \in C_B, \bar{k} \in C_R$, then constraint $\sum_{j \in S} o_{kj} + \sum_{j \in T} o_{\bar{k}j} \geq 1$, for $S = A_k$ and $T = A_{\bar{k}}$, is violated.

To get minimal subsets S and T that keeps the inequalities violated, we proceed as follow. For the V_{BR}-disjoint constraints, we start with $S = A_{\bar{k}}$ and $T = \{i\}$. If there is $u \in S$ such that $H[S \backslash \{u\}]$ still reaches i, then we remove u from S. This procedure finishes when S becomes minimal.

For the V_N-disjoint constraints, we start with $S = A_k$ and $T = A_{\bar{k}}$. If there is $u \in S$ or $w \in T$ such that $H[S \backslash \{u\}] \cap H[T] \cap V_N \neq \emptyset$ or $H[S] \cap H[T \backslash \{w\}] \cap V_N \neq \emptyset$, then we remove u from S or w from T, respectively. This procedure finishes when $S \cup T$ becomes minimal. It is worth noting that the second type of constraints only needs to be verified if $H[S] \cap T = H[T] \cap S = \emptyset$; otherwise it is covered by the first type.

The whole algorithm has complexity $O(\max(|C_B|, |C_R|) \cdot n^4)$.

6.4.5 *Lazy constraints separator for convexity constraints*

We also implemented a lazy constraints separation for the convexity constraints (5.23) and (5.35) of ILP2^M and ILP3^M, since there can be a large number of them for a given instance of the 2-*MGC* problem. It is sufficient to enumerate all such convexity constraints and search for the violated ones, which yields a complexity of $O(|C_{BR}| \cdot n^3)$.

6.5 Geodesic classification algorithms

In this section, we describe the methods used to solve the problem and how they were implemented. The algorithm that we developed for each formulation runs a branch-and-bound algorithm ((LAWLER; WOOD, 1966)), which implicitly enumerates all feasible solutions of the problem via a decision tree structure. It also uses a cutting plane algorithm to solve a linear relaxation of the root node, which includes some valid inequalities (cuts) found by separation algorithms, and a lazy constraint approach to find feasible integer solutions. Such a lazy constraint approach greatly reduces the overall running time as shown by the experiments.

For the formulation ILP1, the main steps of our solution method are described in Algorithm 3. Similarly, the main steps of the solution method for the formulation ILP2 are described in Algorithm 4.

Algorithm 3: ILP1 solving algorithm

Data: Graph Instance G.
Result: Optimal solution for the 2-SGC problem.

1 Computation of all D_{hj} sets using Algorithm 1.
2 Initial cutoff: Since a trivial solution is obtained by taking all vertices of a class as outliers, $\min\{|V_B|, |V_R|\}$ is provided as a cutoff.
3 Initial model configuration: All generalized C_4 inequalities (4.10) (with D_{hj} requirement instead of $H[\{h, j\}]$) are included in the initial model by exhaustive enumeration of all pairs of 2-sized subsets. None of the constraints (4.4)-(4.5) are used initially.
4 Partial linear relaxation resolution: At the root node of the branch-and-cut tree, we solve the linear relaxation of the initial model together with generalized 3-path constraints (4.7) and X-swing constraints (4.8) separated as cuts by enumeration.
5 Exact model resolution: Starting from the model obtained in Step 4, we add the integrality constraints (4.3) and solve the integer formulation by adding (4.4)-(4.5) as lazy constraints (with a Lazy Callback procedure of `CPLEX`).

Algorithm 4: ILP2 solving algorithm

Data: Graph Instance G.
Result: Optimal solution for the 2-SGC problem.

1 Computation of all D_{hj} sets using Algorithm 1.
2 Initial cutoff: Since a trivial solution is obtained by taking all vertices of a class as outliers, $\min\{|V_B|, |V_R|\}$ is provided as a cutoff.
3 Initial model configuration: All generalized C_4 inequalities (4.10) (with D_{hj} requirement instead of $H[\{h, j\}]$) are included in the initial model by exhaustive enumeration of all pairs of 2-sized subsets. All constraints (4.14)-(4.16) are used and none of the constraints (4.17) are included initially.
4 Partial linear relaxation resolution: At the root node of the branch-and-cut tree, we solve the linear relaxation of the initial model together with inequalities (4.21) and (4.25) (with D_{hj} requirement instead of $H[\{h, j\}]$) separated as cuts (for more details of the corresponding separation algorithms, see Section 6.4).
5 Exact model resolution: Starting from the model obtained in Step 4, we add the integrality constraints (4.18) and solve the integer formulation by adding (4.17) as lazy constraints (with a Lazy Callback procedure of `CPLEX`).

6.6 Results and analysis

6.6.1 Random instances

The random instances used in the experiments were categorized by number of vertices ($\{50,100,150,200,250\}$), graph density percentage ($\{5,10,20,30,50,70\}$) and initially classified vertices percentage ($\{20,40,60,80\}$). The number of blue and red initially classified vertices is equal. For each combination of number of vertices v, density d and initially classified vertices percentage br, we generated 10 random instances. Therefore, there were 1200 random instances in total, all of them with parameters $L_B = L_R = 1$.

Tables 2-5 show the running time comparison between solving the complete pure integer linear model ILP2 and running Algorithm 4, while Tables 6-9 present a running time comparison between Algorithm 3 and Algorithm 4. The time limit was set to 3600 seconds (“-"means this time limit was exceeded). The information presented is:

- Instance: instance name using the format: number of vertices in G (v), density of G (d), number of groups per class (l) and initially classified vertices percentage (br);
- Diam: diameter of the graph instance;
- Dgmin: minimum degree of the graph instance;
- Dgmax: maximum degree of the graph instance;
- OPT: optimal solution (the minimum number of outliers);
- $T_{ILP1}(s)$: running time of Algorithm 3 in seconds;
- T_{ILP2}-P(s): running time of solving the complete pure integer linear model ILP2 in seconds;
- $T_{ILP2}(s)$: running time of Algorithm 4 in seconds.

We also studied the effect of the valid inequalities used in each algorithm. For the sake of comparison, we tested two other versions of each algorithm, each version obtained by the elimination of Step 3 or Step 4, respectively. The observed results are summarized in Tables 10 and 11. They compare the performance of the three tested versions with respect to a standard implementation where both steps 3 and 4 were not applied.

We could note that inequalities (4.10) were extremely effective: on average, there were $2.6|V|$ ($12|V|$ for ILP2 when (4.25) were added as cuts) inequalities added and they reduced 82% (83% for ILP2) of the running time and 30% (20% for ILP2) of the number of lazy constraints added. These were the most effective valid inequalities that we found. Remember

Table 2 – Pure integer linear model ILP2 and Algorithm 4 running times comparison for random instances with $br = 20$ (in seconds).

Instance	Diam	Dgmin	Dgmax	OPT	$T_{ILP2\text{-}P}(s)$	$T_{ILP2}(s)$
v050-d05-l1-br20	10	1	6	2	0.12	0.03
v050-d10-l1-br20	5	1	9	4	1.04	0.54
v050-d20-l1-br20	3	4	16	4	2.05	0.30
v050-d30-l1-br20	3	7	22	5	1.94	0.27
v050-d50-l1-br20	2	17	31	4	5.06	0.10
v050-d70-l1-br20	2	26	41	3	2.53	0.12
v100-d05-l1-br20	6	1	10	9	16.64	2.84
v100-d10-l1-br20	4	2	17	9	26.38	2.68
v100-d20-l1-br20	3	9	30	10	17.89	1.22
v100-d30-l1-br20	2	17	40	10	30.15	1.92
v100-d50-l1-br20	2	36	61	10	52.80	0.88
v100-d70-l1-br20	2	56	80	9	42.56	0.62
v150-d05-l1-br20	5	1	14	15	167.20	14.81
v150-d10-l1-br20	3	5	24	15	171.92	16.23
v150-d20-l1-br20	3	17	43	15	132.19	6.81
v150-d30-l1-br20	2	29	59	15	341.01	1.38
v150-d50-l1-br20	2	59	90	15	852.08	0.78
v150-d70-l1-br20	2	90	118	15	1031.14	1.18
v200-d05-l1-br20	4	2	19	20	894.77	35.65
v200-d10-l1-br20	3	8	33	20	1014.29	29.30
v200-d20-l1-br20	3	24	57	20	1082.87	5.87
v200-d30-l1-br20	2	42	78	20	-	2.17
v200-d50-l1-br20	2	80	119	20	-	1.15
v200-d70-l1-br20	2	120	154	20	-	1.17
v250-d05-l1-br20	4	2	23	24	-	140.60
v250-d10-l1-br20	3	12	39	25	-	53.95
v250-d20-l1-br20	2	32	67	25	-	8.52
v250-d30-l1-br20	2	52	96	25	-	4.56
v250-d50-l1-br20	2	100	147	25	-	2.38
v250-d70-l1-br20	2	151	193	25	-	2.77
AVERAGE	-	-	-	-	-	11.36

Table 3 – Pure integer linear model ILP2 and Algorithm 4 running times comparison for random instances with $br = 40$ (in seconds).

Instance	Diam	Dgmin	Dgmax	OPT	$T_{ILP2\text{-}P}(s)$	$T_{ILP2}(s)$
v050-d05-l1-br40	12	1	5	7	0.50	0.11
v050-d10-l1-br40	5	1	9	9	2.73	0.52
v050-d20-l1-br40	3	3	16	9	3.35	0.34
v050-d30-l1-br40	3	7	22	10	3.25	0.23
v050-d50-l1-br40	2	15	32	10	3.56	0.15
v050-d70-l1-br40	2	26	41	9	2.99	0.12
v100-d05-l1-br40	6	1	10	19	25.73	3.07
v100-d10-l1-br40	4	3	17	20	31.79	3.44
v100-d20-l1-br40	3	9	30	20	24.11	1.72
v100-d30-l1-br40	2	19	41	20	45.42	0.73
v100-d50-l1-br40	2	36	61	20	73.98	0.37
v100-d70-l1-br40	2	56	80	20	92.51	0.43
v150-d05-l1-br40	5	1	14	29	197.15	9.04
v150-d10-l1-br40	3	5	26	30	214.28	5.95
v150-d20-l1-br40	3	18	43	30	194.89	2.07
v150-d30-l1-br40	2	31	59	30	429.04	0.33
v150-d50-l1-br40	2	57	89	30	1316.79	1.47
v150-d70-l1-br40	2	89	116	30	1842.82	1.68
v200-d05-l1-br40	4	3	18	40	1149.35	27.60
v200-d10-l1-br40	3	9	32	40	1113.57	8.40
v200-d20-l1-br40	3	25	55	40	1481.16	1.30
v200-d30-l1-br40	2	42	76	40	-	2.36
v200-d50-l1-br40	2	80	119	40	-	2.30
v200-d70-l1-br40	2	121	156	40	-	2.83
v250-d05-l1-br40	4	4	22	50	-	106.01
v250-d10-l1-br40	3	13	39	50	-	11.25
v250-d20-l1-br40	2	32	69	50	-	3.37
v250-d30-l1-br40	2	54	95	50	-	2.99
v250-d50-l1-br40	2	101	145	50	-	6.37
v250-d70-l1-br40	2	153	192	50	-	7.38
AVERAGE	-	-	-	-	-	7.13

Table 4 – Pure integer linear model ILP2 and Algorithm 4 running times comparison for random instances with $br = 60$ (in seconds).

Instance	Diam	Dgmin	Dgmax	OPT	T_{ILP2}-P(s)	$T_{ILP2}(s)$
v050-d05-l1-br60	9	1	6	12	0.94	0.42
v050-d10-l1-br60	4	1	9	14	2.78	0.71
v050-d20-l1-br60	3	4	15	15	3.09	0.54
v050-d30-l1-br60	3	8	22	15	2.84	0.38
v050-d50-l1-br60	2	17	32	15	3.27	0.11
v050-d70-l1-br60	2	26	41	15	3.47	0.12
v100-d05-l1-br60	6	1	10	29	27.70	4.21
v100-d10-l1-br60	4	3	17	29	35.49	4.17
v100-d20-l1-br60	3	10	31	30	32.48	0.71
v100-d30-l1-br60	2	18	41	30	43.60	0.33
v100-d50-l1-br60	2	37	62	30	85.46	1.00
v100-d70-l1-br60	2	57	79	30	97.80	1.14
v150-d05-l1-br60	5	1	15	44	260.04	12.22
v150-d10-l1-br60	3	5	25	45	220.65	3.21
v150-d20-l1-br60	3	17	43	45	239.70	0.43
v150-d30-l1-br60	2	30	59	45	557.84	2.58
v150-d50-l1-br60	2	59	91	45	1316.49	2.39
v150-d70-l1-br60	2	88	118	45	2258.76	2.70
v200-d05-l1-br60	4	3	19	60	1255.83	32.64
v200-d10-l1-br60	3	9	32	60	1207.54	3.75
v200-d20-l1-br60	3	24	55	60	1315.32	1.43
v200-d30-l1-br60	2	43	77	60	-	3.15
v200-d50-l1-br60	2	81	119	60	-	8.10
v200-d70-l1-br60	2	122	156	60	-	5.65
v250-d05-l1-br60	4	4	23	75	-	129.13
v250-d10-l1-br60	3	13	39	75	-	6.64
v250-d20-l1-br60	2	32	68	75	-	2.75
v250-d30-l1-br60	2	54	94	75	-	9.91
v250-d50-l1-br60	2	102	147	75	-	14.20
v250-d70-l1-br60	2	153	194	75	-	16.14
AVERAGE	-	-	-	-	-	9.03

Table 5 – Pure integer linear model ILP2 and Algorithm 4 running times comparison for random instances with $br = 80$ (in seconds).

Instance	Diam	Dgmin	Dgmax	OPT	T_{ILP2}-P(s)	$T_{ILP2}(s)$
v050-d05-l1-br80	10	1	6	15	0.92	0.29
v050-d10-l1-br80	5	1	10	19	2.29	1.05
v050-d20-l1-br80	3	4	16	20	2.32	0.61
v050-d30-l1-br80	3	7	21	20	2.37	0.17
v050-d50-l1-br80	2	16	31	20	3.64	0.09
v050-d70-l1-br80	2	27	41	20	4.17	0.13
v100-d05-l1-br80	6	1	10	39	34.15	4.92
v100-d10-l1-br80	4	3	18	40	39.14	1.09
v100-d20-l1-br80	3	10	29	40	28.95	0.39
v100-d30-l1-br80	2	19	42	40	64.82	0.70
v100-d50-l1-br80	2	37	62	40	79.26	1.41
v100-d70-l1-br80	2	57	80	40	85.24	0.82
v150-d05-l1-br80	5	1	14	59	318.11	5.39
v150-d10-l1-br80	3	6	25	60	200.29	0.60
v150-d20-l1-br80	3	17	43	60	250.78	1.12
v150-d30-l1-br80	2	30	59	60	647.25	2.18
v150-d50-l1-br80	2	59	90	60	1762.73	3.31
v150-d70-l1-br80	2	89	119	60	2425.06	4.41
v200-d05-l1-br80	4	2	19	79	1037.29	13.65
v200-d10-l1-br80	3	9	33	80	1320.49	1.28
v200-d20-l1-br80	2	25	56	80	1739.68	2.93
v200-d30-l1-br80	2	43	77	80	-	9.76
v200-d50-l1-br80	2	80	118	80	-	13.43
v200-d70-l1-br80	2	120	156	80	-	18.47
v250-d05-l1-br80	4	4	22	100	-	9.48
v250-d10-l1-br80	3	13	38	100	-	2.73
v250-d20-l1-br80	2	32	67	100	-	6.51
v250-d30-l1-br80	2	56	95	100	-	17.12
v250-d50-l1-br80	2	102	146	100	-	40.88
v250-d70-l1-br80	2	154	193	100	-	54.93
AVERAGE	-	-	-	-	-	7.33

Table 6 – Algorithm 3 and Algorithm 4 running times comparison for random instances with $br = 20$ (in seconds).

Instance	Diam	Dgmin	Dgmax	OPT	$T_{ILP1}(s)$	$T_{ILP2}(s)$
v050-d05-l1-br20	10	1	6	2	0.01	0.03
v050-d10-l1-br20	5	1	9	4	0.08	0.54
v050-d20-l1-br20	3	4	16	4	0.12	0.30
v050-d30-l1-br20	3	7	22	5	0.10	0.27
v050-d50-l1-br20	2	17	31	4	0.05	0.10
v050-d70-l1-br20	2	26	41	3	0.01	0.12
v100-d05-l1-br20	6	1	10	9	1.14	2.84
v100-d10-l1-br20	4	2	17	9	1.34	2.68
v100-d20-l1-br20	3	9	30	10	1.08	1.22
v100-d30-l1-br20	2	17	40	10	0.96	1.92
v100-d50-l1-br20	2	36	61	10	0.22	0.88
v100-d70-l1-br20	2	56	80	9	0.02	0.62
v150-d05-l1-br20	5	1	14	15	7.96	14.81
v150-d10-l1-br20	3	5	24	15	5.86	16.23
v150-d20-l1-br20	3	17	43	15	2.11	6.81
v150-d30-l1-br20	2	29	59	15	0.56	1.38
v150-d50-l1-br20	2	59	90	15	0.02	0.78
v150-d70-l1-br20	2	90	118	15	0.02	1.18
v200-d05-l1-br20	4	2	19	20	17.55	35.65
v200-d10-l1-br20	3	8	33	20	13.14	29.30
v200-d20-l1-br20	3	24	57	20	3.35	5.87
v200-d30-l1-br20	2	42	78	20	1.06	2.17
v200-d50-l1-br20	2	80	119	20	0.02	1.15
v200-d70-l1-br20	2	120	154	20	0.02	1.17
v250-d05-l1-br20	4	2	23	24	41.13	140.60
v250-d10-l1-br20	3	12	39	25	27.56	53.95
v250-d20-l1-br20	2	32	67	25	2.82	8.52
v250-d30-l1-br20	2	52	96	25	0.44	4.56
v250-d50-l1-br20	2	100	147	25	0.05	2.38
v250-d70-l1-br20	2	151	193	25	0.06	2.77
AVERAGE	-	-	-	-	4.29	11.36

Table 7 – Algorithm 3 and Algorithm 4 running times comparison for random instances with $br = 40$ (in seconds).

Instance	Diam	Dgmin	Dgmax	OPT	$T_{ILP1}(s)$	$T_{ILP2}(s)$
v050-d05-l1-br40	12	1	5	7	0.05	0.11
v050-d10-l1-br40	5	1	9	9	0.63	0.52
v050-d20-l1-br40	3	3	16	9	0.58	0.34
v050-d30-l1-br40	3	7	22	10	0.47	0.23
v050-d50-l1-br40	2	15	32	10	0.08	0.15
v050-d70-l1-br40	2	26	41	9	0.02	0.12
v100-d05-l1-br40	6	1	10	19	4.17	3.07
v100-d10-l1-br40	4	3	17	20	5.05	3.44
v100-d20-l1-br40	3	9	30	20	2.46	1.72
v100-d30-l1-br40	2	19	41	20	0.29	0.73
v100-d50-l1-br40	2	36	61	20	0.01	0.37
v100-d70-l1-br40	2	56	80	20	0.02	0.43
v150-d05-l1-br40	5	1	14	29	14.33	9.04
v150-d10-l1-br40	3	5	26	30	3.82	5.95
v150-d20-l1-br40	3	18	43	30	0.64	2.07
v150-d30-l1-br40	2	31	59	30	0.04	0.33
v150-d50-l1-br40	2	57	89	30	0.09	1.47
v150-d70-l1-br40	2	89	116	30	0.12	1.68
v200-d05-l1-br40	4	3	18	40	23.17	27.60
v200-d10-l1-br40	3	9	32	40	4.63	8.40
v200-d20-l1-br40	3	25	55	40	0.30	1.30
v200-d30-l1-br40	2	42	76	40	0.11	2.36
v200-d50-l1-br40	2	80	119	40	0.16	2.30
v200-d70-l1-br40	2	121	156	40	0.22	2.83
v250-d05-l1-br40	4	4	22	50	52.86	106.01
v250-d10-l1-br40	3	13	39	50	7.32	11.25
v250-d20-l1-br40	2	32	69	50	0.12	3.37
v250-d30-l1-br40	2	54	95	50	0.14	2.99
v250-d50-l1-br40	2	101	145	50	0.40	6.37
v250-d70-l1-br40	2	153	192	50	0.57	7.38
AVERAGE	-	-	-	-	4.10	7.13

Table 8 – Algorithm 3 and Algorithm 4 running times comparison for random instances with $br = 60$ (in seconds).

Instance	Diam	Dgmin	Dgmax	OPT	$T_{ILP1}(s)$	$T_{ILP2}(s)$
v050-d05-l1-br60	9	1	6	12	0.25	0.42
v050-d10-l1-br60	4	1	9	14	1.19	0.71
v050-d20-l1-br60	3	4	15	15	0.83	0.54
v050-d30-l1-br60	3	8	22	15	0.33	0.38
v050-d50-l1-br60	2	17	32	15	0.02	0.11
v050-d70-l1-br60	2	26	41	15	0.01	0.12
v100-d05-l1-br60	6	1	10	29	3.02	4.21
v100-d10-l1-br60	4	3	17	29	2.11	4.17
v100-d20-l1-br60	3	10	31	30	0.52	0.71
v100-d30-l1-br60	2	18	41	30	0.03	0.33
v100-d50-l1-br60	2	37	62	30	0.11	1.00
v100-d70-l1-br60	2	57	79	30	0.13	1.14
v150-d05-l1-br60	5	1	15	44	17.66	12.22
v150-d10-l1-br60	3	5	25	45	1.47	3.21
v150-d20-l1-br60	3	17	43	45	0.07	0.43
v150-d30-l1-br60	2	30	59	45	0.18	2.58
v150-d50-l1-br60	2	59	91	45	0.24	2.39
v150-d70-l1-br60	2	88	118	45	0.31	2.70
v200-d05-l1-br60	4	3	19	60	14.53	32.64
v200-d10-l1-br60	3	9	32	60	0.46	3.75
v200-d20-l1-br60	3	24	55	60	0.28	1.43
v200-d30-l1-br60	2	43	77	60	0.25	3.15
v200-d50-l1-br60	2	81	119	60	0.90	8.10
v200-d70-l1-br60	2	122	156	60	1.19	5.65
v250-d05-l1-br60	4	4	23	75	9.12	129.13
v250-d10-l1-br60	3	13	39	75	0.42	6.64
v250-d20-l1-br60	2	32	68	75	0.28	2.75
v250-d30-l1-br60	2	54	94	75	0.84	9.91
v250-d50-l1-br60	2	102	147	75	2.70	14.20
v250-d70-l1-br60	2	153	194	75	3.89	16.14
AVERAGE	-	-	-	-	2.11	9.03

Table 9 – Algorithm 3 and Algorithm 4 running times comparison for random instances with $br = 80$ (in seconds).

Instance	Diam	Dgmin	Dgmax	OPT	$T_{ILP1}(s)$	$T_{ILP2}(s)$
v050-d05-l1-br80	10	1	6	15	0.31	0.29
v050-d10-l1-br80	5	1	10	19	1.49	1.05
v050-d20-l1-br80	3	4	16	20	1.07	0.61
v050-d30-l1-br80	3	7	21	20	0.16	0.17
v050-d50-l1-br80	2	16	31	20	0.01	0.09
v050-d70-l1-br80	2	27	41	20	0.01	0.13
v100-d05-l1-br80	6	1	10	39	4.28	4.92
v100-d10-l1-br80	4	3	18	40	0.29	1.09
v100-d20-l1-br80	3	10	29	40	0.05	0.39
v100-d30-l1-br80	2	19	42	40	0.08	0.70
v100-d50-l1-br80	2	37	62	40	0.14	1.41
v100-d70-l1-br80	2	57	80	40	0.19	0.82
v150-d05-l1-br80	5	1	14	59	2.67	5.39
v150-d10-l1-br80	3	6	25	60	0.10	0.60
v150-d20-l1-br80	3	17	43	60	0.26	1.12
v150-d30-l1-br80	2	30	59	60	0.24	2.18
v150-d50-l1-br80	2	59	90	60	0.88	3.31
v150-d70-l1-br80	2	89	119	60	1.12	4.41
v200-d05-l1-br80	4	2	19	79	2.96	13.65
v200-d10-l1-br80	3	9	33	80	0.25	1.28
v200-d20-l1-br80	2	25	56	80	0.33	2.93
v200-d30-l1-br80	2	43	77	80	1.04	9.76
v200-d50-l1-br80	2	80	118	80	3.96	13.43
v200-d70-l1-br80	2	120	156	80	5.66	18.47
v250-d05-l1-br80	4	4	22	100	1.60	9.48
v250-d10-l1-br80	3	13	38	100	0.63	2.73
v250-d20-l1-br80	2	32	67	100	0.92	6.51
v250-d30-l1-br80	2	56	95	100	3.14	17.12
v250-d50-l1-br80	2	102	146	100	11.89	40.88
v250-d70-l1-br80	2	154	193	100	16.62	54.93
AVERAGE	-	-	-	-	2.08	7.33

Algorithm 3	N. of constraints (4.7)-(4.8)	N. of inequalities (4.10)	Lazy Const. Reduction	Time Reduction
Step 3 only	0	$2.6\|V\|$	30%	82%
Step 4 only	$14\|V\|$	0	92%	73%
Step 3 and Step 4	$2.9\|V\|$	$2.6\|V\|$	88%	85%

Table 10 – Effect of the valid inequalities for ILP1 for random instances.

Algorithm 4	N. of inequalities (4.21)	N. of inequalities (4.10), (4.25)	Lazy Const. Reduction	Time Reduction
Step 3 only	0	$12\|V\|$	20%	83%
Step 4 only, with (4.21)	$8\|V\|$	0	85%	11%
Step 3 and Step 4	$5\|V\|$	$17\|V\|$	79%	87%

Table 11 – Effect of the valid inequalities for ILP2 for random instances.

that inequalities (4.10) were proved to be facet-defining for the polytope associated with ILP1 (for more details, see Section 4.2.2) and for the polytope associated with ILP2 (if the generalized C_4 is actually an C_4). However, the generalized C_4 constraints (4.25) included as cuts in the root node of the branch-and-cut tree showed only a bit improvement of the linear relaxation lower bound in Algorithm 4. It is important to note that there were no good results when all the generalized C_4 constraints (4.25) were included in the initial model of ILP2.

Constraints (4.7)-(4.8), added when solving the root node in Step 4 of Algorithm 3, were very effective as well. Adding them as cuts was much better than including all of them in the initial model. On average, there were $14|V|$ constraints added and they reduced 73% of the running time and 92% of the number of lazy constraints added. On the other hand, the generalized convexity inequalities (4.21), added when solving the root node in Step 4 of Algorithm 4, were not very effective in reducing the running time for the random instances, although reducing 85% of the number of lazy constraints added. On average, there were $8|V|$ constraints added and they reduced only 11% of the running time. However, we anticipate that, for the realistic instances, the running time reduction was 70%, and they showed to be very useful. Since there are many of these inequalities, including all of them in the initial model did not give good results, as expected. The combination of all cited inequalities (i.e, including Steps 3 and 4 in both algorithms) showed an overall running time reduction of 85% for ILP1 and 87% for ILP2, and it drastically reduced the number of lazy constraints added (88% for ILP1 and 79% for ILP2). Moreover, the addition of the generalized C_4 inequalities yield a very good reduction of inequalities (4.7), (4.8) and (4.21) added in the root node.

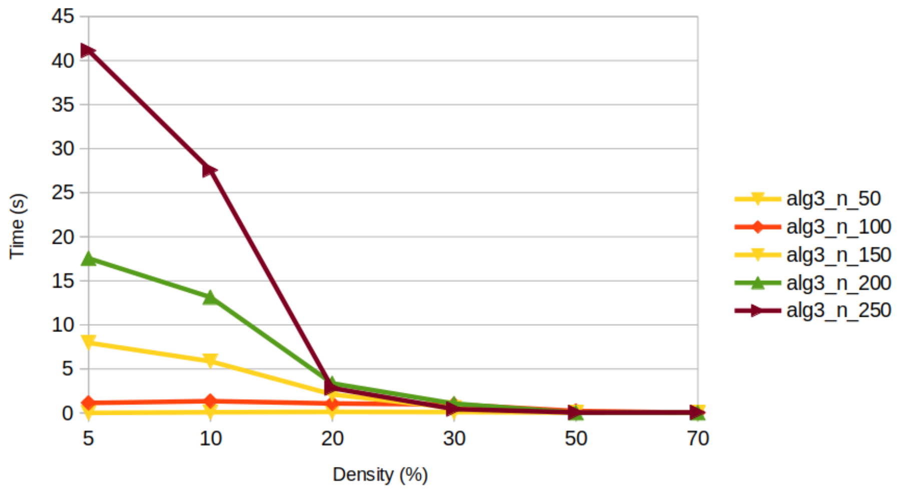
Star tree inequalities (4.11), (4.19) and generalized walk inequalities (4.24) did not reduce the running time, so we did not use them in the final version of the branch-and-cut algorithms.

Regarding the lazy constraints scheme presented in Section 6.4.4, its application in Step 5 was fundamental to reduce the running time (with respect to an implementation with all constraints (4.4)-(4.5) added to the initial model). Actually, it is impractical to solve the problem without the lazy constraints scheme since the number of constraints (4.4)-(4.5) is potentially exponential. Considering all random instances, the average running time of Algorithm 3 was about few seconds, so it is shown to be very good even for medium size instances.

In Tables 2-5, we can see how the use of the valid inequalities and the lazy constraints scheme in Algorithm 4 greatly reduce the running time to solve ILP2. For almost all instances with the number of vertices greater than or equal to 200, solving the complete pure formulation exceeded the time limit. The main information that we can retrieve from these experiments is that the lazy constraints scheme for the convexity constraints works very well, even for medium instances, as already noticed.

Tables 6-9 also show the efficiency of the generalized C_4 inequalities and the lazy constraints scheme presented in Section 6.4.4 for formulation ILP1. For almost all instances, Algorithm 3 had beaten Algorithm 4 in running time, yielding an overall running time $T_{ILP1}(s)$ smaller than $T_{ILP2}(s)$. The few cases in which Algorithm 4 produced better results lied on instances of a low number of vertices and low density. Besides, Algorithm 3 has been shown to be extremely efficient for dense instances, even for medium instances, because in such cases the size of $S \cup T$ for the model constraints is generally smaller, which can reduce the number of constraints. Then, for the overall results, $T_{ILP1}(s)$ was better than $T_{ILP2}(s)$ in 111 instance configurations from the total of 120, which is 92% of all instances. Given that, Algorithm 3 seems to be better than Algorithm 4 for random instances.

Figure 29 – Running time versus density: Algorithm 3 and $br = 20$.



Figures 29, 31, 33 and 35 show the running times of Algorithm 3 as a function of the graph density, whereas Figures 30, 32, 34 and 36 show these results for Algorithm 4. There is a graphic for each value of br, where the value of n varies with $\{50, 100, 150, 200, 250\}$. These results show evidences that Algorithm 3 works extremely well, especially for dense, medium sized instances. Besides, it can be seen that, for $br \leq 40$ and for sufficient large n, the running times of Algorithm 3 and Algorithm 4 decrease exponentially as the density increases. Moreover, in general, the instances with density between 5% and 20% or $br = 80\%$ were the hardest to solve.

Figure 30 – Running time versus density: Algorithm 4 and $br = 20$.

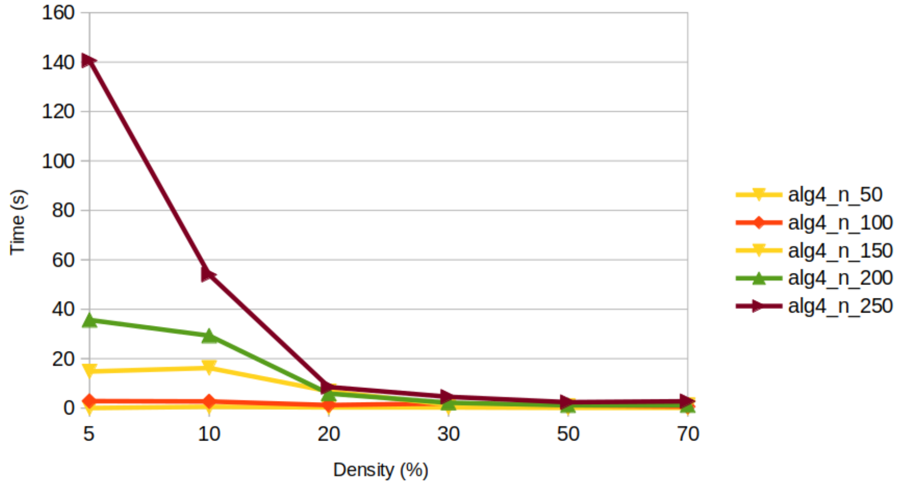


Figure 31 – Running time versus density: Algorithm 3 and $br = 40$.

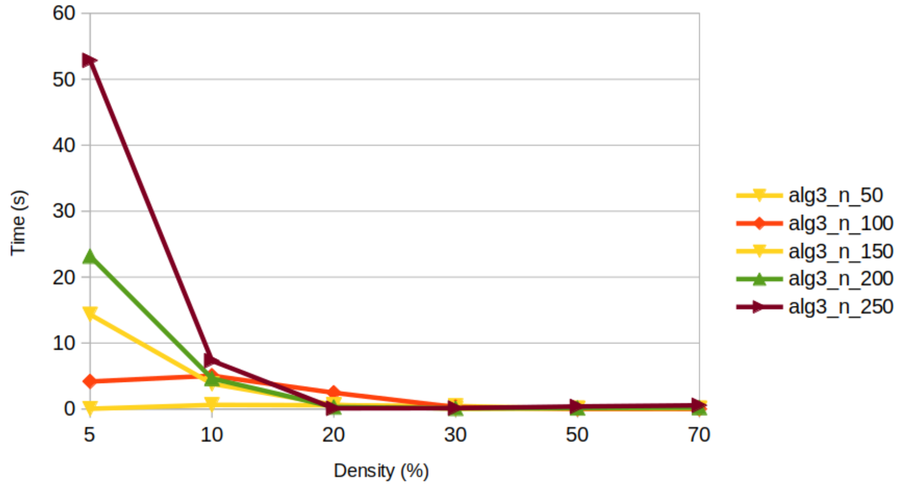


Figure 32 – Running time versus density: Algorithm 4 and $br = 40$.

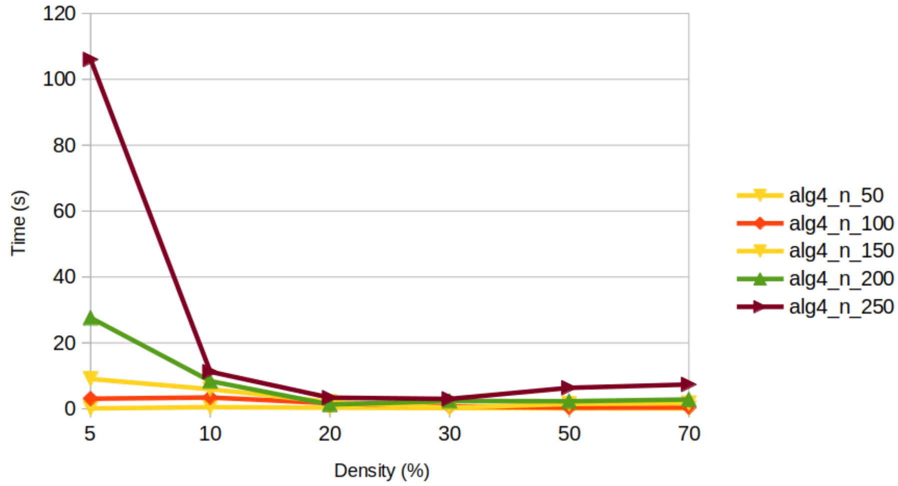


Figure 33 – Running time versus density: Algorithm 3 and $br = 60$.

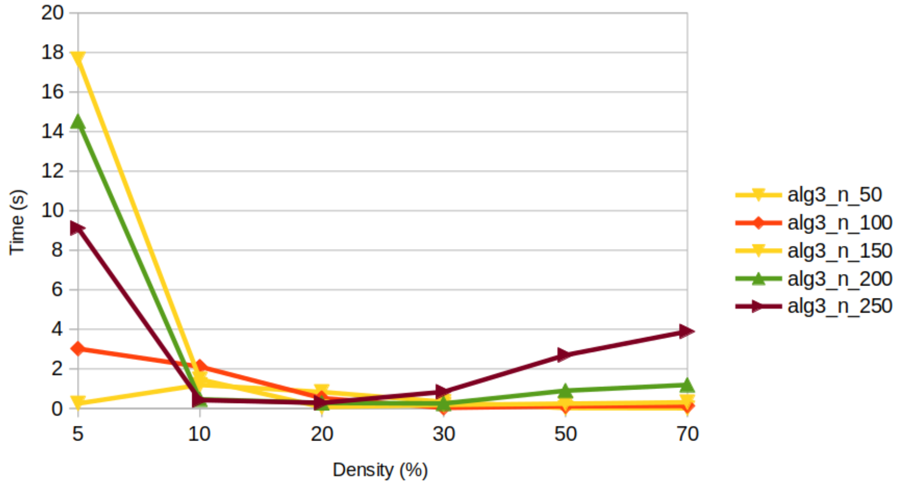


Figure 34 – Running time versus density: Algorithm 4 and $br = 60$.

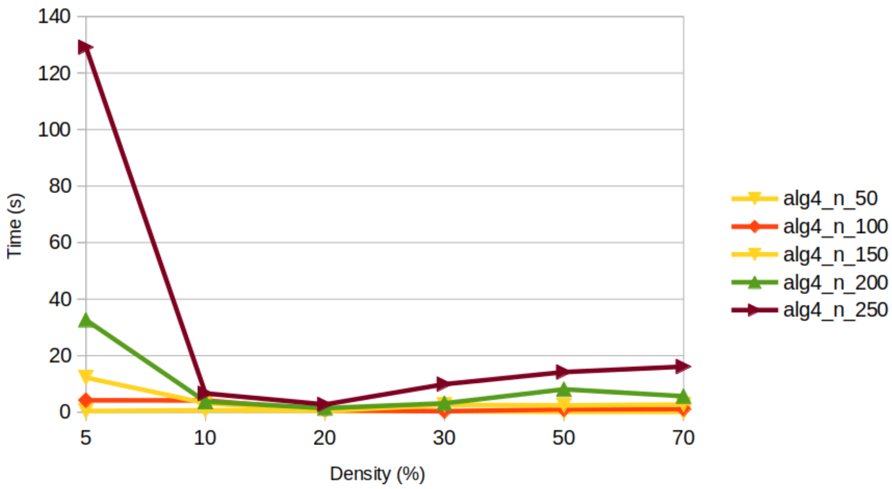
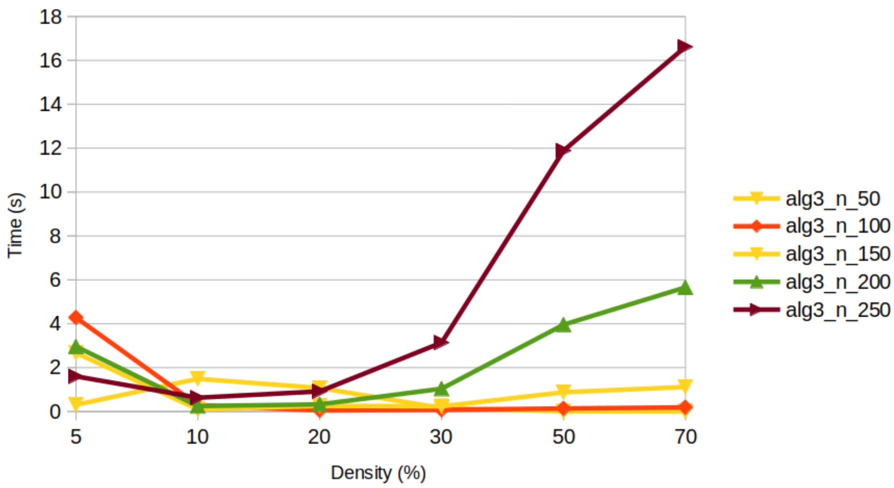


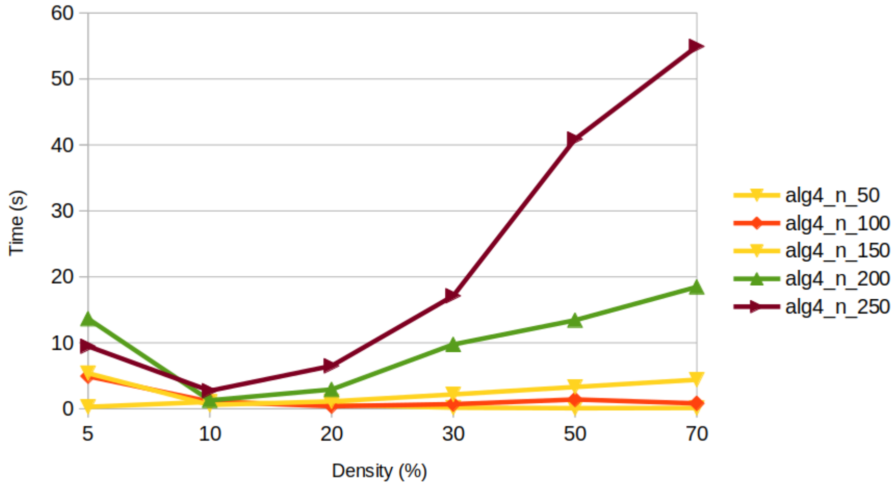
Figure 35 – Running time versus density: Algorithm 3 and $br = 80$.



6.6.2 *Realistic and synthetic instances*

To test the developed algorithms for realistic applications, we performed experiments using instances derived from two realistic datasets, namely Parkinson's disease ((LITTLE *et al.*,

Figure 36 – Running time versus density: Algorithm 4 and $br = 80$.



2007)) and cardiac Single Proton Emission Computed Tomography (SPECT) images ((KURGAN *et al.*, 2001)), both available at https://archive.ics.uci.edu/ml/datasets.html. The datasets come from instances of the Euclidean version of the classification problem, in which each point represents the information of a patient to be used to predict new diagnostics. So, as a way to evaluate the accuracy of class prediction of our algorithms, we also run the classic *SVM* and the *MLP* Euclidean classification algorithms for these datasets (see Sections 6.6.3 and 6.6.4).

For each of these datasets, we derived 10 associated instances for the 2-*SGC* problem in the following way. We constructed the associated classification graph using the transformation suggested by (ZAKI; JR, 2014), where each point becomes a vertex. Then, we randomly chose 20% (or 30%) of the vertices to become unclassified (they form the validation set). These were the vertices (points in the Euclidean version) used to test the efficiency and accuracy of the geodesic and Euclidean classification algorithms by comparing their predefined classes, available in the datasets, with the predicted classes from the solutions returned by Algorithm 3, Algorithm 4, *SVM* and the *MLP* algorithms.

The detailed explanation of the dataset transformation into a graph is presented next. Each vertex i of the graph represents a point x_i of the dataset. To create the edges, we first compute the pairwise similarity between the points using the Gaussian kernel function given by:

$$a_{ij} = exp\{\frac{-||x_i - x_j||^2}{2}\},$$

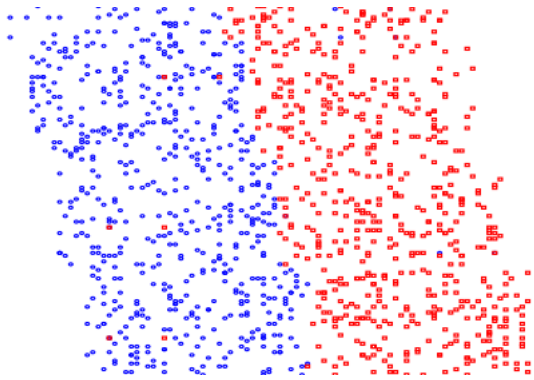
where $||x_i - x_j||^2 = \sum_k (x_i(k) - x_j(k))^2$ is the square of the Euclidean distance between points x_i and x_j. Then, each pair (x_i, x_j) has a similarity weight a_{ij} corresponding to the similarity value between x_i and x_j. Next, for each vertex i, we compute the top q nearest neighbors in terms of the similarity value, given as

$$N_q(i) = \{j \in V \mid j \neq i, a_{ij} \geq a_{iq}\},$$

where a_{iq} represents the similarity value between i and its q-th most similar neighbor (we used $q = 10\%$ of $|V|$). Then, an edge is added between vertices i and j if, and only if, $j \in N_q(i)$ and $i \in N_q(j)$. Finally, if the resulting graph is disconnected, we add the top q most similar (i.e., highest weighted) edges between each pair of connected components.

We follow a similar approach to compare the results for the synthetic instances. By synthetic instance, we mean a randomly generated instance where the optimum solution is known in advance. As in (BLAUM *et al.*, 2019a), we generated two base synthetic instances for the Euclidean classification problem, as follows. Given the space dimension $d \in \{2,3\}$ and a hyperplane in $\mathbb{R}^d$, we chose one point in each half-space, say b and r, each at a distance 1 of the hyperplane. Point b (resp. r) will work as the center for the blue (resp. red) samples. Then, we consider a d-dimensional hypercube with side 2 centered at each center point; we randomly generate p points uniformly distributed within it. Finally, we generate 5 outliers for each class. We use $p = 50$ and $n = 5$. An illustration of the base instance for $d = 2$ is shown in Figure 37. From each base instance, we derive 10 Euclidean instances by choosing 20% (and also 30%) of the points to become unclassified. The corresponding graph instances are obtained with the same procedure used for Parkinson's and SPECT instances.

Figure 37 – Synthetic instance example for $d = 2$.



6.6.3 SVM *algorithm*

A *Support Vector Machine* (*SVM*) is a discriminative classifier formally defined by a separating hyperplane. In other words, given labeled training data (supervised learning), the algorithm outputs an optimal hyperplane which categorizes new examples. In a two dimensional space, this hyperplane is a line dividing a plane into two parts where in each class lays in one of the sides.

In our experiments, we used the *linearSVC* algorithm implementation of the *SVM*, found at http://scikit-learn.org/stable/modules/svm.html, as an Euclidean classification algorithm. It uses a linear kernel, which means that it tries to separate the dataset linearly. Thus, it is appropriated for comparison with our geodesic classification algorithms, since we used parameters $LB = LR = 1$ in our experiments.

6.6.4 **Neural networks** ***algorithm MLP***

As another Euclidean classification algorithm, we used a *Neural Networks* implementation of a *Multi-Layer Perceptron* (*MLP*) algorithm, which applies *backpropagation*. For more details, see http://scikit-learn.org/stable/modules/neural-networks-supervised.html.

6.6.5 Parkinson's disease instances

In the Parkinson's disease dataset, each point represents the information of biomedical voice measurement of a person that may or not have the Parkinson's disease. The main aim of the data is to discriminate healthy people from those with Parkinson's disease, according to some previously known information (initially classified points).

The dataset is composed of 195 points, each of them containing a person's biomedical voice measurement information divided into 22 attributes. Thus, each point belonging to $\mathbb{R}^{22}$. Among these 195 points, 48 points are related to healthy people and 147 to people with Parkinson's disease.

6.6.6 SPECT heart data instances

The cardiac dataset describes diagnosing of cardiac Single Proton Emission Computed Tomography (*SPECT*) images. Each of the patients is classified into two categories (classes): normal and abnormal. The database consists of 267 *SPECT* image sets (patients) that were processed to extract features that summarize the original *SPECT* images. As a result, 44 continuous feature patterns were created for each patient, along with the diagnosis status (normal or abnormal). This information is used to provide a set of diagnoses for cardiac *SPECT* studies.

Table 12 – Algorithm 3, Algorithm 4, *SVM* and *MLP* comparison for Parkinson's instances with $br = 80\%$.

Instance	Diam	Dgmin	Dgmax	OPT	$T_{ILP1}(s)$	$T_{ILP2}(s)$	$Acu_{GC}(\%)$	$Acu_{SVM}(\%)$	$Acu_{MLP}(\%)$
parkinsons-v195-d5-l1-br80-1	10	1	18	39	68.52	6.21	79.49%	87.18%	79.49%
parkinsons-v195-d5-l1-br80-2	10	1	18	38	89.23	6.18	74.36%	74.36%	71.79%
parkinsons-v195-d5-l1-br80-3	10	1	18	33	57.42	1.10	64.10%	64.10%	74.36%
parkinsons-v195-d5-l1-br80-4	10	1	18	34	88.23	3.65	64.10%	64.10%	76.92%
parkinsons-v195-d5-l1-br80-5	10	1	18	38	36.81	2.95	76.92%	76.92%	82.05%
parkinsons-v195-d5-l1-br80-6	10	1	18	39	151.17	49.64	76.92%	76.92%	76.92%
parkinsons-v195-d5-l1-br80-7	10	1	18	38	92.19	48.60	74.36%	79.49%	71.79%
parkinsons-v195-d5-l1-br80-8	10	1	18	33	53.75	4.11	61.54%	61.54%	69.23%
parkinsons-v195-d5-l1-br80-9	10	1	18	39	169.02	4.84	79.49%	23.08%	79.49%
parkinsons-v195-d5-l1-br80-10	10	1	18	38	40.36	5.00	79.49%	79.49%	71.79%
AVERAGE	-	-	-	-	87.64	13.23	73.08%	68.72%	75.38%

6.6.7 Analysis of the realistic instances experiments

Tables 12 (instances with 80% of initially classified vertices) and 13 (instances with 70% of initially classified vertices) present the running time comparison between Algorithm 3 and Algorithm 4, along with the prediction accuracy comparison between the 2-*SGC*, the *SVM* and the *MLP* approaches for the Parkinson's instances. Similarly, Tables 14 (instances with 80% of initially classified vertices) and 15 (instances with 70% of initially classified vertices) show the corresponding comparison for the *SPECT* instances. The information presented is:

- Instance: instance name using the format: application name, number of vertices in G (v), density of G (d), number of groups per class (l), percentage of initially classified vertices (br) and instance number;
- Diam: diameter of the graph instance;
- Dgmin: minimum degree of the graph instance;
- Dgmax: maximum degree of the graph instance;
- OPT: optimal solution (the minimum number of outliers);
- $T_{ILP1}(s)$: running time of Algorithm 3 in seconds;
- $T_{ILP2}(s)$: running time of Algorithm 4 in seconds;
- $Acu_{GC}(\%)$: percentage of correct class prediction from solutions of Algorithm 3 and Algorithm 4 (number of correct predictions divided by the number of unclassified vertices);
- $Acu_{SVM}(\%)$: percentage of correct class prediction from solutions of the *SVM* algorithm;
- $Acu_{MLP}(\%)$: percentage of correct class prediction from solutions of the *MLP* algorithm.

A first remark concerns the high number of outliers in the optimal solution. Remember that we are using linear separation and these realistic instances are more unlikely to have such a property.

Regarding the Parkinson's disease instances, we can see that 2-*SGC* obtained the

Table 13 – Algorithm 3, Algorithm 4, *SVM* and *MLP* comparison for Parkinson's instances with $br = 70\%$.

Instance	Diam	Dgmin	Dgmax	*OPT*	$T_{ILP1}(s)$	$T_{ILP2}(s)$	$Acu_{GC}(\%)$	$Acu_{SVM}(\%)$	$Acu_{MLP}(\%)$
parkinsons-v195-d5-l1-br70-1	10	1	18	40	488.66	28.34	86.21%	63.79%	86.21%
parkinsons-v195-d5-l1-br70-2	10	1	18	29	11.69	1.84	68.97%	68.97%	74.14%
parkinsons-v195-d5-l1-br70-3	10	1	18	34	99.07	2.73	77.59%	77.59%	81.03%
parkinsons-v195-d5-l1-br70-4	10	1	18	31	27.10	6.36	70.69%	70.69%	70.69%
parkinsons-v195-d5-l1-br70-5	10	1	18	33	21.48	13.83	74.14%	25.86%	72.41%
parkinsons-v195-d5-l1-br70-6	10	1	18	34	306.83	20.98	75.86%	75.86%	72.41%
parkinsons-v195-d5-l1-br70-7	10	1	18	34	24.94	2.52	79.31%	77.59%	77.59%
parkinsons-v195-d5-l1-br70-8	10	1	18	34	66.23	7.20	75.86%	24.14%	75.86%
parkinsons-v195-d5-l1-br70-9	10	1	18	38	288.10	41.48	82.76%	84.48%	81.03%
parkinsons-v195-d5-l1-br70-10	10	1	18	34	58.73	46.92	75.86%	82.76%	75.86%
AVERAGE	-	-	-	-	139.28	17.22	76.72%	65.17%	76.72%

Table 14 – Algorithm 3, Algorithm 4, *SVM* and *MLP* comparison for *SPECT* instances with $br = 80\%$.

Instance	Diam	Dgmin	Dgmax	*OPT*	$T_{ILP1}(s)$	$T_{ILP2}(s)$	$Acu_{GC}(\%)$	$Acu_{SVM}(\%)$	$Acu_{MLP}(\%)$
spectf-v267-d5-l1-br80-1	8	1	36	47	0.20	1.08	86.79%	79.25%	86.79%
spectf-v267-d5-l1-br80-2	8	1	36	42	0.14	0.72	77.36%	83.02%	77.36%
spectf-v267-d5-l1-br80-3	8	1	36	47	0.29	0.76	86.79%	71.70%	83.02%
spectf-v267-d5-l1-br80-4	8	1	36	43	0.15	0.62	79.25%	75.47%	79.25%
spectf-v267-d5-l1-br80-5	8	1	36	42	0.13	0.93	77.36%	49.06%	77.36%
spectf-v267-d5-l1-br80-6	8	1	36	42	0.15	0.74	77.36%	79.25%	77.36%
spectf-v267-d5-l1-br80-7	8	1	36	46	0.32	4.63	83.02%	81.13%	83.02%
spectf-v267-d5-l1-br80-8	8	1	36	41	0.17	0.76	75.47%	73.58%	75.47%
spectf-v267-d5-l1-br80-9	8	1	36	43	0.12	0.93	77.36%	75.47%	77.36%
spectf-v267-d5-l1-br80-10	8	1	36	41	0.13	0.61	75.47%	67.92%	75.47%
AVERAGE	-	-	-	-	0.18	1.18	79.62%	73.58%	79.25%

Table 15 – Algorithm 3, Algorithm 4, *SVM* and *MLP* comparison for *SPECT* instances with $br = 70\%$.

Instance	Diam	Dgmin	Dgmax	*OPT*	$T_{ILP1}(s)$	$T_{ILP2}(s)$	$Acu_{GC}(\%)$	$Acu_{SVM}(\%)$	$Acu_{MLP}(\%)$
spectf-v267-d5-l1-br70-1	8	1	36	41	0.17	0.66	83.75%	72.50%	83.75%
spectf-v267-d5-l1-br70-2	8	1	36	37	0.26	0.61	78.75%	73.75%	78.75%
spectf-v267-d5-l1-br70-3	8	1	36	38	0.16	0.94	80.00%	75.00%	80.00%
spectf-v267-d5-l1-br70-4	8	1	36	35	0.17	0.70	76.25%	67.50%	75.00%
spectf-v267-d5-l1-br70-5	8	1	36	40	0.37	0.76	81.25%	41.25%	81.25%
spectf-v267-d5-l1-br70-6	8	1	36	39	0.23	0.54	80.00%	62.50%	80.00%
spectf-v267-d5-l1-br70-7	8	1	36	37	0.20	0.67	78.75%	51.25%	78.75%
spectf-v267-d5-l1-br70-8	8	1	36	41	0.22	0.88	83.75%	60.00%	83.75%
spectf-v267-d5-l1-br70-9	8	1	36	46	0.43	0.61	88.75%	87.50%	86.25%
spectf-v267-d5-l1-br70-10	8	1	36	38	0.15	0.55	78.75%	76.25%	72.50%
AVERAGE	-	-	-	-	0.24	0.69	81.00%	66.75%	80.00%

best accuracy in 10 of them, while 9 and 11 were the corresponding scores for *SVM* and *MLP*, respectively, from the total of 20 instances. However, on average, *SVM* got the worst accuracy, due to the poor performance in some instances. 2-*SGC* and *MLP* obtained similar average accuracy with a slight advantage to the latter. For these instances, we could observe higher running times for Algorithm 3 in comparison with the random instances. This was possibly due to the larger diameters and lower maximum degrees of the input graphs. For this reason, Algorithm 4 got the best running time for all Parkinson's disease instances in these experiments.

For the SPECT instances, we can observe that the running times of our geodesic classification algorithms were much smaller than those for the Parkinson's instances. Note that

Table 16 – Algorithm 3, Algorithm 4, *SVM* and *MLP* comparison for synthetic instances with dimension 2 and $br = 80\%$.

Instance	Diam	Dgmin	Dgmax	*OPT*	$T_{ILP1}(s)$	$T_{ILP2}(s)$	$Acu_{GC}(\%)$	$Acu_{SVM}(\%)$	$Acu_{MLP}(\%)$
syntheticDim2-v104-d7-l1-br80-1	17	1	9	13	0.31	0.77	100.00%	90.00%	50.00%
syntheticDim2-v104-d7-l1-br80-2	17	1	9	13	0.83	1.71	90.00%	95.00%	50.00%
syntheticDim2-v104-d7-l1-br80-3	17	1	9	8	0.39	0.50	70.00%	65.00%	30.00%
syntheticDim2-v104-d7-l1-br80-4	17	1	9	11	0.40	0.34	90.00%	70.00%	40.00%
syntheticDim2-v104-d7-l1-br80-5	17	1	9	8	0.22	0.38	75.00%	55.00%	30.00%
syntheticDim2-v104-d7-l1-br80-6	17	1	9	9	0.32	0.34	85.00%	80.00%	45.00%
syntheticDim2-v104-d7-l1-br80-7	17	1	9	11	0.59	0.56	95.00%	75.00%	45.00%
syntheticDim2-v104-d7-l1-br80-8	17	1	9	12	0.50	0.69	95.00%	85.00%	40.00%
syntheticDim2-v104-d7-l1-br80-9	17	1	9	11	0.11	0.49	95.00%	90.00%	45.00%
syntheticDim2-v104-d7-l1-br80-10	17	1	9	10	0.16	0.24	90.00%	85.00%	40.00%
AVERAGE	-	-	-	-	0.38	0.60	88.50%	79.00%	41.50%

Table 17 – Algorithm 3, Algorithm 4, *SVM* and *MLP* comparison for synthetic instances with dimension 2 and $br = 70\%$.

Instance	Diam	Dgmin	Dgmax	*OPT*	$T_{ILP1}(s)$	$T_{ILP2}(s)$	$Acu_{GC}(\%)$	$Acu_{SVM}(\%)$	$Acu_{MLP}(\%)$
syntheticDim2-v104-d7-l1-br70-1	17	1	9	13	2.95	2.77	93.55%	87.10%	45.16%
syntheticDim2-v104-d7-l1-br70-2	17	1	9	12	0.82	0.51	90.32%	87.10%	48.39%
syntheticDim2-v104-d7-l1-br70-3	17	1	9	11	1.05	1.57	90.32%	67.74%	48.39%
syntheticDim2-v104-d7-l1-br70-4	17	1	9	11	0.38	0.74	96.77%	90.32%	38.71%
syntheticDim2-v104-d7-l1-br70-5	17	1	9	13	0.73	0.60	96.77%	96.77%	48.39%
syntheticDim2-v104-d7-l1-br70-6	17	1	9	13	0.88	1.88	90.32%	80.65%	48.39%
syntheticDim2-v104-d7-l1-br70-7	17	1	9	12	0.76	0.60	96.77%	93.55%	45.16%
syntheticDim2-v104-d7-l1-br70-8	17	1	9	10	0.26	0.71	87.10%	90.32%	35.48%
syntheticDim2-v104-d7-l1-br70-9	17	1	9	11	0.37	0.53	96.77%	87.10%	45.16%
syntheticDim2-v104-d7-l1-br70-10	17	1	9	13	4.60	3.51	90.32%	70.97%	38.71%
AVERAGE	-	-	-	-	1.28	1.34	92.90%	85.16%	44.19%

in the SPECT graphs the diameters are lower and the maximum degree are higher. Regarding accuracy, the 2-*SGC* approach presented the best accuracy in 18 SPECT instances, while *SVM* and *MLP* did it in 2 and 14 instances, respectively. On average, 2-*SGC* got also the best accuracy, slightly better than the one by *MLP*.

Overall, the results show that the accuracy of the 2-*MGC* approach was the best for 28 instances, while *SVM* was the best for only 11 and *MLP* for 25, from the total of 40 instances.

6.6.8 Analysis of the synthetic instances experiments

Tables 16 (instances with 80% of initially classified vertices) and 17 (instances with 70% of initially classified vertices) present the running time comparison between Algorithm 3 and Algorithm 4, along with the prediction accuracy comparison between the 2-*SGC*, the *SVM* and the *MLP* approaches for the synthetic instances with dimension 2, while Tables 18 (instances with 80% of initially classified vertices) and 19 (instances with 70% of initially classified vertices) show the corresponding comparison for the instances with dimension 3.

Regarding the synthetic instances with dimension 2, 2-*SGC* obtained the best accuracy in 18 of them, while 3 and 0 were the corresponding scores for *SVM* and *MLP*, respectively, from the total of 20 instances. On average, 2-*SGC* got the best result and *MLP* got a bad result.

Table 18 – Algorithm 3, Algorithm 4, *SVM* and *MLP* comparison for synthetic instances with dimension 3 and $br = 80\%$.

Instance	Diam	Dgmin	Dgmax	OPT	$T_{ILP1}(s)$	$T_{ILP2}(s)$	$Acu_{GC}(\%)$	$Acu_{SVM}(\%)$	$Acu_{MLP}(\%)$
syntheticDim3-v96-d6-l1-br80-1	10	2	8	16	0.27	0.58	94.74%	94.74%	47.37%
syntheticDim3-v96-d6-l1-br80-2	10	2	8	17	0.31	0.83	100.00%	100.00%	42.11%
syntheticDim3-v96-d6-l1-br80-3	10	2	8	14	0.19	0.33	73.68%	94.74%	47.37%
syntheticDim3-v96-d6-l1-br80-4	10	2	8	17	0.21	0.41	89.47%	94.74%	42.11%
syntheticDim3-v96-d6-l1-br80-5	10	2	8	15	0.12	0.31	100.00%	100.00%	63.16%
syntheticDim3-v96-d6-l1-br80-6	10	2	8	16	0.26	0.76	84.21%	94.74%	68.42%
syntheticDim3-v96-d6-l1-br80-7	10	2	8	14	0.41	0.55	73.68%	89.47%	21.05%
syntheticDim3-v96-d6-l1-br80-8	10	2	8	15	0.32	0.33	94.74%	94.74%	47.37%
syntheticDim3-v96-d6-l1-br80-9	10	2	8	15	0.23	0.41	84.21%	94.74%	63.16%
syntheticDim3-v96-d6-l1-br80-10	10	2	8	15	0.13	0.34	89.47%	89.47%	47.37%
AVERAGE	-	-	-	-	0.25	0.49	88.42%	94.74%	48.95%

Table 19 – Algorithm 3, Algorithm 4, *SVM* and *MLP* comparison for synthetic instances with dimension 3 and $br = 70\%$.

Instance	Diam	Dgmin	Dgmax	OPT	$T_{ILP1}(s)$	$T_{ILP2}(s)$	$Acu_{GC}(\%)$	$Acu_{SVM}(\%)$	$Acu_{MLP}(\%)$
syntheticDim3-v96-d6-l1-br70-1	10	2	8	12	0.11	0.12	78.57%	96.43%	50.00%
syntheticDim3-v96-d6-l1-br70-2	10	2	8	16	0.28	0.15	78.57%	100.00%	64.29%
syntheticDim3-v96-d6-l1-br70-3	10	2	8	11	0.15	0.38	92.86%	89.29%	57.14%
syntheticDim3-v96-d6-l1-br70-4	10	2	8	13	0.11	0.38	85.71%	96.43%	50.00%
syntheticDim3-v96-d6-l1-br70-5	10	2	8	11	0.08	0.18	78.57%	85.71%	39.29%
syntheticDim3-v96-d6-l1-br70-6	10	2	8	13	0.10	0.34	82.14%	96.43%	53.57%
syntheticDim3-v96-d6-l1-br70-7	10	2	8	14	0.26	1.23	89.29%	96.43%	28.57%
syntheticDim3-v96-d6-l1-br70-8	10	2	8	15	0.40	2.78	85.71%	96.43%	32.14%
syntheticDim3-v96-d6-l1-br70-9	10	2	8	12	0.13	0.20	82.14%	92.86%	53.57%
syntheticDim3-v96-d6-l1-br70-10	10	2	8	13	0.05	0.27	78.57%	96.43%	57.14%
AVERAGE	-	-	-	-	0.17	0.60	83.21%	94.64%	48.57%

In most instances, Algorithm 3 has beaten Algorithm 4 in running time, but the difference was very small.

On the other hand, for the synthetic instances with dimension 3, 2-*SGC* obtained the best accuracy in 6 of them, while 19 and 0 were the corresponding scores for *SVM* and *MLP*, respectively, from the total of 20 instances. On average, *SVM* got the best result and, again, *MLP* got a very bad result. Algorithm 3 has beaten Algorithm 4 in running time in all but one instance, showing that Algorithm 3 was better.

Overall, the results show that the accuracy of the 2-*SGC* approach was the best for 24 instances, while *SVM* was the best for 22 and *MLP* for none of them, from the total of 40 instances. Therefore, for these experiments, the performance of 2-*SGC* showed to be similar to that of *SVM*.

7 CONCLUDING REMARKS

In this work, we defined two versions of the geodesic classification problem on graphs (2-class single-group geodesic classification and 2-class multi-group geodesic classification problems) as the analog of the Euclidean classification problem. These new problems present pure combinatorial optimization aspects and appear as an intersection of a graph convexity problem and the well-known set covering problem. Their applications arise in the fields of data mining and statistics, which have been increasingly studied in recent years.

We proposed three integer programming formulations for these new combinatorial optimization problems. As the main focus of this work, we studied the polyhedra associated with these formulations for the single-group and the multi-group cases, giving some valid inequalities and facet-defining conditions. We also established conditions to transform valid and facet-defining inequalities from the single-group case to the multi-group case, making a parallel between the polyhedra of both cases. In order to run computational experiments to validate the accuracy of the geodesic classification approach and the efficiency of the proposed valid inequalities, we developed a branch and cut algorithm to solve, exactly, the integer formulations for the single-group case. An interesting point of the theoretical results is that one of the families of facet-defining inequalities, namely the generalized C_4 inequalities, is related to a structure of mutual convex combinations that is not possible to appear in the Euclidean space. Moreover, they showed to be extremely useful, as stated by the experiments.

The results of the computational experiments also show that the proposed solution methods are very promising since the branch-and-cut algorithms for the integer formulations proved to be very efficient (in running time and accuracy), even for medium-sized instances. The algorithm for the set covering formulation $ILP1$ was the best one for the most of the tested instances. It is important to remark that the proposed lazy constraints scheme and a cutting plane algorithm were fundamental to reduce the running time.

We validated the accuracy of the geodesic convexity approach by comparing the prediction provided by the proposed algorithms with two of the most used approaches for the Euclidean convexity classification problem, namely *SVM* and *MLP*. The prediction accuracy of the geodesic approach showed to be stable and as good as such classic linear separation algorithms for the multidimensional space. Therefore, it seems that the analogy performed to transform the Euclidean convexity method into a geodesic convexity method on graphs was successful.

An interesting idea for the use of the classification graph would be the use of weighted edges determined by how often the vertices receive the same class or the same group index given by various resolution methods for the Euclidean classification problem. Then, the weighted edges would try to simulate the underlying pattern of the samples better, and a new classification approach in such a graph could be even more accurate.

As future works, we intend to carry on a deeper polyhedral study and try to eliminate the symmetries in the formulations to improve the performance of the branch-and-cut algorithms. Although we know some cases in which the geodesic classification problem can be solved in polynomial time, it seems to be an NP-hard problem in general due to its similarities with other NP-hard problems that involve geodesic convexity on graphs. But it remains an open question whether that problem does belong to the NP-hard class set. It would also be interesting to study other variants of the classification problem, such as the multi-class variant, where more than 2 classes are used to classify the samples. For this variant, the formulations proposed for the 2-class version can be easily adapted to characterize its solution set. Thus, most of the techniques and methods presented in this work can be also applied to the multi-class version of the geodesic classification problem.

REFERENCES

ARAÚJO, P. H. M.; CAMPÊLO, M.; CORRÊA, R. C.; LABBÉ, M. The geodesic classification problem on graphs. **Latin & American Algorithms, Graphs and Optimization Symposium**, [*s.n.*], Belo Horizonte, Brasil, 2019.

ARTHANARI, T. S.; DODGE, Y. **Mathematical Programming in Statistics**. New York: Wiley-Interscience, 1993. ISBN 978-0-471-59212-9.

ARTIGAS, D.; DANTAS, S.; DOURADO, M. C.; SZWARCFITER, J. L. Partitioning a graph into convex sets. **Discrete Mathematics**, [*S.l.: s.n.*], v. 311, n. 17, p. 1968–1977, 2011.

BALAS, E. Projection with a minimal system of inequalities. **Computational Optimization and Applications**, [*S.l.: s.n.*], v. 10, n. 2, p. 189–193, May 1998. ISSN 1573-2894.

BALAS, E.; NG, S. M. On the set covering polytope: I. all the facets with coefficients in {0, 1, 2}. **Mathematical Programming**, [*S.l.: s.n.*], v. 43, n. 1, p. 57–69, Jan 1989. ISSN 1436-4646.

BALAS, E.; OOSTEN, M. On the dimension of projected polyhedra. **Discrete Applied Mathematics**, [*S.l.: s.n.*], v. 87, n. 1, p. 1 – 9, 1998. ISSN 0166-218X.

BERTSIMAS, D.; SHIODA, R. Classification and regression via integer optimization. **Operations Research**, [*S.l.: s.n.*], v. 55, n. 2, p. 252–271, 2007.

BLAUM, M.; CORRÊA, R. C.; MARENCO, J.; KOCH, I.; MYDLARZ, M. An integer programming approach for the 2-class single-group classification problem. **Latin & American Algorithms, Graphs and Optimization Symposium**, [*s.n.*], Belo Horizonte, Brasil, 2019.

BLAUM, M.; CORRÊA, R. C.; MARENCO, J.; KOCH, I.; MYDLARZ, M. A set covering approach for the 2-class classification problem. Working paper. 2019.

BONDY, J. A.; MURTY, U. S. R. **Graph Theory**. [*S.l.*]: Springer, 2008.

BUZATU, R.; CATARANCIUC, S. Convex graph covers. **The Computer Science Journal of Moldova**, [*S.l.: s.n.*], v. 23, n. 3, p. 251–269, 2015.

CáCERES, J.; HERNANDO, C.; MORA, M.; PELAYO, I. M.; PUERTAS, M. L.; SEARA, C. On geodetic sets formed by boundary vertices. **Discrete Math.**, Elsevier Science Publishers B. V., Amsterdam, The Netherlands, The Netherlands, v. 306, n. 2, p. 188–198, fev. 2006. ISSN 0012-365X.

CAMPÊLO, M.; CAMPOS, V. A.; CORRÊA, R. C. On the asymmetric representatives formulation for the vertex coloring problem. **Discrete Applied Mathematics**, [*S.l.: s.n.*], v. 156, n. 7, p. 1097 – 1111, 2008. ISSN 0166-218X. GRACO 2005.

CARRIZOSA, E.; MORALES, D. R. A mixed integer optimisation model for data classification. **Computers & Operations Research**, Pergamon, [*S.l.*], v. 40, n. 1, p. 150–165, 2013.

CHANGAT, M.; MATHEW, J. On triangle path convexity in graphs. **Discrete Mathematics**, [*S.l.: s.n.*], v. 206, n. 1, p. 91 – 95, 1999. ISSN 0012-365X.

CORNUÉJOLS, G.; SASSANO, A. On the 0,1 facets of the set covering polytope. **Mathematical Programming**, [*S.l.: s.n.*], v. 43, p. 45–55, 01 1989.

CORRÊA, R. C.; DONNE, D. D.; MARENCO, J. On the combinatorics of the 2-class classification problem. **Discrete Optimization**, [*S.l.: s.n.*], v. 31, p. 40–55, 2019. ISSN 1572-5286.

CORTES, C.; VAPNIK, V. Support-vector networks. In: **Machine Learning**. [S.l.: s.n.], 1995. p. 273–297.

DOURADO, M. C.; GIMBEL, J. G.; KRATOCHVÍL, J.; PROTTI, F.; SZWARCFITER, J. L. On the computation of the hull number of a graph. **Discrete Math.**, Elsevier Science Publishers B. V., Amsterdam, The Netherlands, The Netherlands, v. 309, n. 18, p. 5668–5674, set. 2009. ISSN 0012-365X.

DOURADO, M. C.; GIMBEL, J. G.; KRATOCHVíL, J.; PROTTI, F.; SZWARCFITER, J. L. On the computation of the hull number of a graph. **Discrete Mathematics**, [*S.l.: s.n.*], v. 309, n. 18, p. 5668 – 5674, 2009. ISSN 0012-365X. Combinatorics 2006, A Meeting in Celebration of Pavol Hell's 60th Birthday (May 1-5, 2006).

DOURADO, M. C.; PROTTI, F.; RAUTENBACH, D.; SZWARCFITER, J. L. Some remarks on the geodetic number of a graph. **Discrete Math.**, Elsevier Science Publishers B. V., Amsterdam, The Netherlands, The Netherlands, v. 310, n. 4, p. 832–837, fev. 2010. ISSN 0012-365X.

EDELMAN, P.; JAMISON, R. The theory of convex geometries. [*S.l.: s.n.*], v. 19, p. 247–270, 12 1985.

FARBER, M.; JAMISON, R. E. Convexity in graphs and hypergraphs. **SIAM Journal on Algebraic Discrete Methods**, [*S.l.: s.n.*], v. 7, n. 3, p. 433–444, 1986.

FERREIRA, C. E.; WAKABAYASHI, Y. **Combinatória Poliédrica e Planos-de-Corte Faciais**. Campinas, São Paulo, Brasil: UNICAMP, 1996.

FREED, N.; GLOVER, F. Evaluating alternative linear programming models to solve the two group discriminant problem. **Decision Sciences**, [*S.l.: s.n.*], v. 17, n. 2, p. 151–162, 2007.

HARARY, F.; NIEMINEN, J. Convexity in graphs. **J. Differential Geom.**, Lehigh University, [*S.l.*], v. 16, n. 2, p. 185–190, 1981.

HONG, L.; DAVISON, B. D. Empirical study of topic modeling in twitter. In: **Proceedings of the First Workshop on Social Media Analytics**. New York, NY, USA: ACM, 2010. (SOMA '10), p. 80–88. ISBN 978-1-4503-0217-3.

KARP, R. M. Reducibility among combinatorial problems. In: ____. **Complexity of Computer Computations**. Boston, MA: Springer US, 1972. p. 85–103. ISBN 978-1-4684-2001-2.

KURGAN, L. A.; CIOS, K. J.; TADEUSIEWICZ, R.; OGIELA, M.; GOODENDAY, L. S. Knowledge discovery approach to automated cardiac spect diagnosis. **Artificial intelligence in medicine**, [*S.l.: s.n.*], v. 23, n. 2, p. 149–169, Oct 2001. ISSN 0933-3657.

LAWLER, E. L.; WOOD, D. E. Branch-and-bound methods: A survey. **Oper. Res.**, INFORMS, Institute for Operations Research and the Management Sciences (INFORMS), Linthicum, Maryland, USA, v. 14, n. 4, p. 699–719, ago. 1966. ISSN 0030-364X.

LIMA, K. **Recoloração Convexa de Caminhos**. Tese (Doutorado) — PhD thesis, Instituto de Matemática e Estatística da Universidade de São Paulo, São Paulo, Brasil, 2011.

LITTLE, M. A.; MCSHARRY, P. E.; ROBERTS, S. J.; COSTELLO, D. A.; MOROZ, I. M. Exploiting nonlinear recurrence and fractal scaling properties for voice disorder detection. **BioMedical Engineering OnLine**, [*S.l.: s.n.*], v. 6, n. 1, p. 23, Jun 2007. ISSN 1475-925X.

MASKOOKI, A. Improving the efficiency of a mixed integer linear programming based approach for multi-class classification problem. **Comput. Ind. Eng.**, Pergamon Press, Inc., Tarrytown, NY, USA, v. 66, n. 2, p. 383–388, out. 2013. ISSN 0360-8352.

NEMHAUSER, G. L.; WOLSEY, L. A. **Integer and combinatorial optimization**. [*S.l.*]: Wiley, 1988. (Wiley interscience series in discrete mathematics and optimization). ISBN 978-0-471-82819-8.

PARDALOS, P. M.; HANSEN, P. **Data mining and mathematical programming**. American Mathematical Society, Providence, RI: [*s.n.*], 2008. v. 45.

PELAYO, I. M. **Geodesic Convexity in Graphs**. New York: Springer-Verlag, 2013.

ROCKAFELLAR, R. **Convex Analysis**. [*S.l.*]: Princeton University Press, 1997. (Convex Analysis).

SÁNCHEZ-GARCÍA, M.; SOBRÓN, M. I.; VITORIANO, B. On the set covering polytope:facets with coefficients in $\{0, 1, 2, 3\}$. **Annals of Operations Research**, [*S.l.: s.n.*], v. 81, p. 343–356, Jun 1998. ISSN 1572-9338.

SCHRIJVER, A. **Theory of Linear and Integer Programming**. New York, NY, USA: John Wiley & Sons, Inc., 1986. ISBN 0-471-90854-1.

SHIODA, R. **Integer optimization in data mining; Data mining via integer optimization**. Tese (Doutorado) — PhD thesis, Massachusetts Institute of Technology, Sloan School of Management, Operations Research Center, 2003.

SUN, M. A mixed integer programming model for multiple-class discriminant analysis. **International Journal of Information Technology and Decision Making**, [*S.l.: s.n.*], v. 10, n. 4, p. 589–612, 2011.

UNEY, F.; TURKAY, M. A mixed-integer programming approach to multi-class data classification problem. **European Journal of Operational Research**, [*S.l.: s.n.*], v. 173, n. 3, p. 910–920, 2006.

VANDERBECK, F.; WOLSEY, L. A. Reformulation and decomposition of integer programs. In: ____. **50 Years of Integer Programming 1958-2008: From the Early Years to the State-of-the-Art**. Berlin, Heidelberg: Springer Berlin Heidelberg, 2010. p. 431–502. ISBN 978-3-540-68279-0.

XU, G.; PAPAGEORGIOU, L. G. A mixed integer optimisation model for data classification. **Comput. Ind. Eng.**, Pergamon Press, Inc., Tarrytown, NY, USA, v. 56, n. 4, p. 1205–1215, maio 2009. ISSN 0360-8352.

ZAKI, M. J.; JR, W. M. **Data Mining and Analysis: Fundamental Concepts and Algorithms**. New York, NY, USA: Cambridge University Press, 2014. ISBN 0521766338, 9780521766333.