



**HAL**  
open science

## Maximizing minimum cycle bases intersection

Ylène Aboufath, Dimitri Watel, Marc-Antoine Weisser, Thierry Mautor,  
Dominique Barth

► **To cite this version:**

Ylène Aboufath, Dimitri Watel, Marc-Antoine Weisser, Thierry Mautor, Dominique Barth. Maximizing minimum cycle bases intersection. 2022. hal-03851365

**HAL Id: hal-03851365**

**<https://hal.science/hal-03851365v1>**

Preprint submitted on 14 Nov 2022

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Maximizing Minimum Cycle Bases Intersection

Ylène Aboufath<sup>1</sup>, Dimitri Watel<sup>2,3</sup>, Marc-Antoine Weisser<sup>4</sup>, Thierry Mautor<sup>1</sup>, and Dominique Barth<sup>1</sup>

<sup>1</sup> DAVID, Université Versailles Saint-Quentin-En-Yvelines, France  
ylene.aboufath@uvsq.fr, dominique.barth@uvsq.fr, thierry.mautor@uvsq.fr

<sup>2</sup> SAMOVAR, Telecom SudParis, France

<sup>3</sup> ENSIIE, Evry, France

dimitri.watel@ensiie.fr

<sup>4</sup> LISN, Centrale Supélec, France

marc-antoine.weisser@centralesupelec.fr

**Abstract.** In this paper we consider the problem of, given a set of graphs with the same set of vertices, choosing a minimum cycle basis for each of these graphs such that the intersection of these bases has maximal size. We first prove this problem to be NP-complete, then prove its polynomiality for the special case of two graphs and study both the hardness of approximation and the parameterized complexity.

## Introduction

The cycle bases are a compact representation of the cycles in a graph, and finding a minimum cycle basis (minimizing the total weight of cycles in the base) can be done in polynomial time [1,12]. We consider an extension of this problem: given a set of graphs with the same set of vertices, find for each graph a minimum cycle basis such that the size of their intersection is maximum.

The motivation for studying this problem is related to chemoinformatics and bioinformatics, especially to study molecular dynamics trajectories. A molecule consists of atoms (i.e., vertices) and covalent/hydrogen bonds between atoms (i.e., edges). The hydrogen bonds can evolve over time [5]. Such an evolution is typically studied using notions of similarities or editing distance between molecular graphs resulting from the trajectory; this is also the case for RNA graphs [11,19]. Considering the paradigm indicating that the structure of a molecule is the consequence of the interaction of its smallest rings (or cycles), a possible approach consists in studying the similarity of molecular graphs of a trajectory through the similarity of their graphs of cycles obtained by choosing cycle bases [8,9,13]. Therefore, the prerequisite is associating a minimum cycle basis to each molecular graph in the molecular trajectory with the maximum global intersection. This is the specific problem we are addressing here.

Minimum cycle bases are already used in chemistry to represent the molecular structure [3,15,17]. Concerning the similarity measure between molecular graphs, some works propose using cycles as a local structure to approximate an edit distance between different graphs [13]. A first study has been carried out concerning the similarity of graphs of cycles of molecular graphs to compare their structure [4], but to our knowledge, no study has been devoted to searching for bases of similar cycles.

**Our contributions.** Given a set of  $k$  graphs with the same set of vertices, this paper introduces the problem of finding for each graph a minimum cycle basis such that the size of their intersection is maximum. This paper shows the following results:

1. The problem is NP-complete when  $k \geq 3$ .
2. The problem is polynomial when  $k = 2$ .
3. There is a polynomial algorithm with performance guarantee  $\frac{1}{k}$  for the associated maximization problem, and this ratio is the best possible.

4. The problem is in XP and is W[1]-hard, parameterized by the size of the solution.

**Method and organization of the paper** This paper introduces the problem MAX-MCBI that, given a set of  $k$  graphs with the same set of vertices, searches a minimum cycle basis for each of these graphs such that the intersection of these bases has maximal size. We denote by MCBI the decision problem associated. We introduce the problem and classic definitions of graphs and minimum cycle bases in Section 1. One can notice similarities between MCBI and the matroid intersection problem. Hence, in the same section, we detail similarities and differences between those two problems. We prove the NP-completeness when  $k \geq 3$  in Section 2. For that, we define a gadget in Section 2.1 and a reduction from the Hamiltonian path problem in Section 2.2. Polynomial case of  $k = 2$ , hardness of approximability and parameterized complexity of the maximization problem are studied in Section 3. Finally, in Section 3.4, we propose an implementation of two functions recurrently needed in our algorithms.

## 1 Minimum Cycle Bases and Matroids

Brief reminders of the definitions of cycles and their bases, as initially presented in [2], are first given. Then the problem of minimum cycle bases intersection is presented.

We consider a general definition of a cycle in a graph  $G$  as any subgraph in which each vertex has an even degree. The sum of two cycles,  $c_1 \oplus c_2$ , is the subgraph containing only the set of edges present in one of the two cycles. This general definition of cycles with the sum operation defines a vectorial space, the space of cycles  $\mathcal{C}$  of a graph. The dimension of the cycle space  $\mathcal{C}$  of a graph  $G = (V, E)$  is  $\mu(G) = |E| - |V| + x$  where  $x$  is the number of connected components in  $G$ . A cycle basis of a graph  $G$  is a set  $\mathcal{B}$  of cycles that spans the cycle space  $\mathcal{C}$  of the graph  $G$ . By span, we refer to the linear span of linear algebra that can be applied here on cycle bases due to their definition in the vectorial space  $\mathcal{C}$ .

The weight of a cycle is its number of edges, denoted for a cycle  $c$  by  $\omega(c)$ . The weight of a cycle basis is the sum of the weights of the cycles that compose it. Therefore, a minimum cycle basis is a cycle basis in the space  $\mathcal{C}$  that minimizes its weight. We denote the set of minimum cycle bases of a graph  $G$  by  $\mathcal{MCB}(G)$ . There are polynomial time algorithms to find a minimum cycle basis [1,12], and these algorithms also work in the weighted setting, where each edge has a non-negative weight associated with it.

According to the general definition of cycles, an elementary cycle is a connected subgraph in which each vertex has degree two. Moreover, we specify the following property: the cycles of a minimum basis are always elementary. Indeed, they can generate any cycle of  $\mathcal{C}$  and remain shorter. Hence, each cycle can be expressed as a  $\{0, 1\}$  incidence vector, i.e., a vector in which each factor represents an edge  $e$  and values 1 if  $e$  is in the cycle and 0 otherwise. Thus, the sum of two cycles is the sum of their vectors modulo 2.

To represent the links between cycles and minimum cycle bases we consider the following notation: given  $\mathcal{B}$  a minimum cycle basis (or any subset of a cycle basis), let  $\lambda_{\mathcal{B}} : \mathcal{B} \times \mathcal{C} \rightarrow \{0, 1\}$  be an application such that if a cycle  $c \in \mathcal{B}$  takes part in the generation of a cycle  $d$  we have  $\lambda_{\mathcal{B}}(c, d) = 1$ , otherwise  $\lambda_{\mathcal{B}}(c, d) = 0$ .

Let us now introduce the decision problem we are interested in.

*Problem 1 (Minimum Cycle Bases Intersection, MCBI).* Given a set of  $k$  graphs  $G_i$  with the same set of vertices and  $\gamma \in \mathbb{N}$ , does it exist a set  $\{\mathcal{B}_1, \dots, \mathcal{B}_k\}$  with  $\mathcal{B}_i \in \mathcal{MCB}(G_i)$  such that  $|\bigcap_{i=1}^k \mathcal{B}_i| \geq \gamma$ .

The maximization problem associated with the decision problem MCBI that searches a set of minimum cycle bases that maximizes the intersection size is denoted by MAX-MCBI. One can notice that this problem is closely related to the matroid intersection problem that we now recall.

A matroid is a pair  $(S, I)$  where  $S$  is a universal set called the ground set, and  $I$  is a family of  $2^{|S|}$  subset called the independent subsets of  $S$ . A matroid has the following properties: (i) if  $f \in I$  then any subset of

$f$  is independent; (ii) if  $f_1$  and  $f_2$  are independents with  $|f_1| < |f_2|$  then  $f_1$  can be completed by an element  $e$  of  $f_2$  such that  $f_1 \cup e \in I$ .

*Problem 2 (Matroid intersection problem).* Given a set of  $k$  matroids  $(S, I_1), (S, I_2), \dots, (S, I_k)$  with the same ground set and  $\gamma \in \mathbb{N}$ , does it exist a subset  $S' \subseteq S$  with  $S' \subseteq \bigcap_{i=1}^k I_i$  such that  $|S'| \geq \gamma$ .

Any vectorial space  $E$  where the independent sets are the linear independent sets of  $E$  defines a matroid. Thus, if  $S$  is the set of cycles in a graph  $G$  and  $I$  are the set of linear independent cycles, then  $(S, I)$  is a matroid. Consequently, the problem MCBI seems to be a subproblem of the matroid intersection problem. The problem of matroid intersection is known to be polynomial in the case of two matroids and NP-complete for more than three matroids [18]. However, we cannot easily apply the results of this intersection problem to MCBI. Our consideration of only minimum cycle bases is one of the main reasons that makes it difficult. Actually, relaxing the problem MCBI by considering cycle bases instead of *minimum* cycle bases makes the problem polynomial. The exponential number of cycles in a graph is another difficulty. Indeed, the complexity of the known polynomial algorithms for matroids intersection is by the size of the ground set, which has, in our case, an exponential size.

We now express common lemmas on minimum cycle bases: Lemmas 1, 2 and 3 respectively present how to construct a basis from another one, how a cycle can only be generated by a shorter cycle, and how the number of cycles of a particular size is limited in a basis.

**Lemma 1.** *Given  $\mathcal{B}$  a cycle basis of a graph  $G$  with two cycles  $c_1 \notin \mathcal{B}$  and  $c_2 \in \mathcal{B}$ , if  $\lambda_{\mathcal{B}}(c_2, c_1) = 1$  then  $\mathcal{B}' = (\mathcal{B} \setminus \{c_2\}) \cup \{c_1\}$  is a cycle basis of  $G$ .*

*Proof.* As  $c_1 = c_2 \oplus \left( \bigoplus_{c \in \mathcal{B} \setminus \{c_2\}} \lambda_{\mathcal{B}}(c, c_1) \cdot c \right)$  then  $c_2 = c_1 \oplus \left( \bigoplus_{c \in \mathcal{B}' \setminus \{c_1\}} \lambda_{\mathcal{B}}(c, c_2) \cdot c \right)$ . Thus  $\mathcal{B}'$  spans every cycle generated by  $\mathcal{B}$ . Given that  $|\mathcal{B}'| = |\mathcal{B}|$  then  $\mathcal{B}'$  is a cycle basis of  $G$ .

**Lemma 2.** *Given a graph  $G$  and  $\mathcal{B} \in \mathcal{MCB}(G)$ , for any cycles  $c_1, c_2$  with  $c_1 \notin \mathcal{B}$  and  $c_2 \in \mathcal{B}$  such that  $\lambda_{\mathcal{B}}(c_2, c_1) = 1$ , we have  $\omega(c_2) \leq \omega(c_1)$ .*

*Proof.* If  $\omega(c_2) > \omega(c_1)$  then the exchange of  $c_2$  with  $c_1$  in  $\mathcal{B}$  gives a basis  $\mathcal{B}'$  (Lemma 1) with a lower weight. So,  $\mathcal{B}$  is not a minimum cycle basis that represents a contradiction.

**Lemma 3.** *Given a graph  $G$  and  $l \in \mathbb{N}$ , for any  $\mathcal{B} \in \mathcal{MCB}(G)$  the subset of cycles  $c \in \mathcal{B}$  with  $\omega(c) \leq l$  spans the same subset of cycles of  $G$ .*

*Proof.* The subset of cycles with a length inferior or equal to  $l$  in a minimum cycle basis  $\mathcal{B}$  is denoted by  $\mathcal{B}(l)$ . Given  $\mathcal{B}, \mathcal{B}' \in \mathcal{MCB}(G)$ , for all  $c$  with  $\omega(c) > l$ , if  $c$  can be generated by  $\mathcal{B}(l)$  then  $c = \bigoplus_{d \in \mathcal{B}(l)} \lambda_{\mathcal{B}}(d, c) \cdot d$ . However, by Lemma 2, every cycle  $d \in \mathcal{B}(l)$  must be spanned by cycles of  $\mathcal{B}'(l)$ ; otherwise,  $d$  should be added to  $\mathcal{B}'$  so that it constitutes a minimum cycle basis. Hence,  $\forall d \in \mathcal{B}(l)$ , we have  $d = \bigoplus_{e \in \mathcal{B}'(l)} \lambda_{\mathcal{B}'}(e, d) \cdot e$ . Thus,  $c$  can also be generated by  $\mathcal{B}'(l)$ .

**Lemma 4.** *Given a graph  $G$  and  $l \in \mathbb{N}$ , for all  $\mathcal{B} \in \mathcal{MCB}(G)$ , the subset of  $\mathcal{B}$  containing all cycles  $c \in \mathcal{B}$  with  $\omega(c) = l$  has the same size.*

*Proof.* This lemma naturally results from Lemma 3. For  $\mathcal{B}, \mathcal{B}' \in \mathcal{MCB}(G)$  as  $\mathcal{B}(l)$  and  $\mathcal{B}'(l)$  span the same set of cycles, if one has a size larger than the other, it cannot be minimum.

## 2 Complexity of MCBI in the general case

This section first introduces a gadget that encodes a method of selecting cycles. The minimum cycle bases of this gadget have the property to contain only one of the given squares. Function GAD takes a set of cycles as an entry and represents this gadget. Then, we give the proof of NP-completeness of problem MCBI through a polynomial reduction to the Hamiltonian path problem.

### 2.1 Encoding a selection of one cycle among $p$ , function GAD

Given  $p$  disjoint cycles  $c_i = (r_i, s_i, t_i, u_i)$  of length 4,  $GAD(c_1, c_2, \dots, c_p)$  is a connected planar embedded graph around those cycles (example in Figure 1). In the following, GAD will be used both to designate the graph in question and also the procedure that builds this graph that is the following:

1. For each pair of consecutive cycles  $c_i, c_{i+1}$  placed side by side on the plane: add two edges  $[s_i, r_{i+1}]$ ,  $[t_i, u_{i+1}]$ , and one node  $w_i$  in between which is connected to  $s_i, r_{i+1}, u_{i+1}$  and  $t_i$ .
2. Let  $e_1, \dots, e_q$  be the edges of the external face with  $q = 4p$ . Add, for each  $1 \leq i \leq q$ , a node  $v_i$  on the external face and a triangle containing  $v_i$  and  $e_i$ . For example, if  $i = 1$  then add  $[u_1, v_1]$  and  $[v_1, r_1]$  to construct the triangle  $u_1, v_1, r_1$  containing  $v_1$  and  $e_1$ .
3. On the external face, for  $1 \leq i < q$  add an edge  $[v_i, v_{i+1}]$  and then, add one edge  $[v_q, v_1]$ .
4. Still on the external face, for each  $3 \leq i \leq q - 1$  add an edge  $[v_1, v_i]$ .

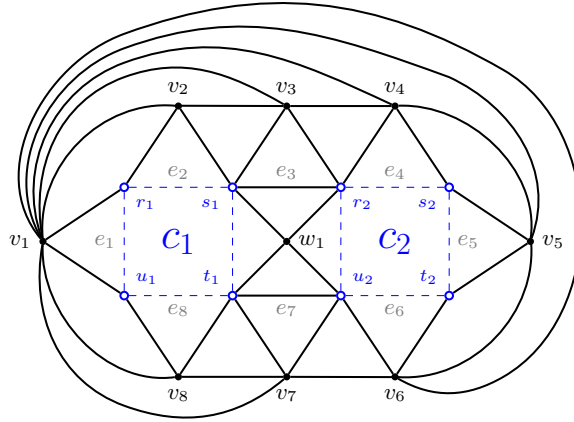


Fig. 1:  $GAD(c_1, c_2)$ ,  $c_1$  and  $c_2$  are represented by dashed blue lines.

In the resulting graph  $GAD(c_1, \dots, c_p)$ , denoted as GAD, we consider only two types of cycles: triangles and squares. Triangles are cycles of length three with at least one vertex not in the given cycles  $c_i$ . Squares are cycles of length four, and, are either one of the given cycles  $c_i$  or are generated by the triangles.

Each triangle is a face (internal or external), but every face is not a triangle. Indeed,  $\mu(GAD)$  cycles constitute the set of inner faces, and that includes the  $p$  squares  $c_i$ . A cycle basis contains  $\mu(GAD)$  linearly independent cycles, and as two faces never have the same set of edges, the set of inner faces is a cycle basis. Hence, we can define any  $\mathcal{B} \in \mathcal{MCB}(GAD)$  by replacing one of the squares  $c_1, c_2, \dots, c_p$  by the outer face

in the set of all the inner faces of GAD. The obtained  $\mathcal{B}$  is a minimum cycle basis because there exists no more independent triangle and thus none to be exchanged with a square to reduce the weight of the basis. From Lemma 4, we know that the number of squares in any minimum cycle basis of the graph is always the same. Consequently, we can express the following property.

*Property 1.* Any  $\mathcal{B} \in \mathcal{MCB}(\text{GAD}(c_1, c_2, \dots, c_p))$  contains  $p - 1$  cycles among the  $p$  given cycles  $c_1, c_2, \dots, c_p$ .

## 2.2 MCBI is NP-complete

We consider the problem that, given a directed graph  $G = (V, A)$  investigates whether there exists a Hamiltonian path between  $s$  and  $t$  in  $G$ . We assume there are no incoming arcs in  $s$  and no outgoing arcs in  $t$ . This problem of the Hamiltonian path is NP-complete [14].

We use a similar proof to the one used for the NP-Completeness of the matroid intersection problem [18]. In that proof, the idea is to define 3 matroids with the same ground set, the set of arcs of  $G$ , but with distinct independent sets: in  $M_1$ , independent arcs cannot form cycles; in  $M_2$ , the independent arcs do not share the same vertex as an origin; and, in  $M_3$ , the independent arcs do not share the same vertex as a destination. Then the maximum common independent set of these three matroids is necessarily an elementary path, i.e., a path passing only once by an arc/a node. The maximum common independent is of size  $|V| - 1$  if and only if the independent represents a hamiltonian path.

We consider an instance  $G = (V, A)$  of the Hamiltonian path problem. We define three graphs  $G_1$ ,  $G_2$  and  $G_3$  that will have the same properties as the matroids  $M_1$ ,  $M_2$  and  $M_3$ . In each graph we add a square  $c_a$  for each arc  $a \in A$ . We will simulate the matroids' independents in the graphs by our function GAD. The three graphs  $G_1$ ,  $G_2$ , and  $G_3$  must have the same vertices. Hence each node that will be added by functions GAD to one graph, will also be added to the two others as a disconnected node.

In  $G_2$ , we consider each node  $v \in V$  one by one. We apply function  $\text{GAD}(c_{(v,w_1)}, c_{(v,w_2)})$  for each pair of successors  $(w_1, w_2)$  of  $v$ . Let us consider  $G_2^v$ , a connected component of  $G_2$  formed by the squares of the outgoing arcs of  $v$ . A minimum cycle basis of  $G_2^v$  will contain every triangle and only one square. Thus, we cannot select two outgoing arcs of  $v$ . Indeed, any minimum cycle basis of a connected component of  $G_2$ , contains every triangle from all the functions GAD and only one square. For example, given a node  $v$  and its three successors  $w_1$ ,  $w_2$  and  $w_3$  represented by the squares  $c_1 = c_{(v,w_1)}$ ,  $c_2 = c_{(v,w_2)}$  and  $c_3 = c_{(v,w_3)}$ . We define  $G_2^v = c_1 \cup c_2 \cup c_3 \cup \text{GAD}(c_1, c_2) \cup \text{GAD}(c_2, c_3) \cup \text{GAD}(c_1, c_3)$ . We denote the minimum cycle basis of  $G_2^v$  by  $\mathcal{B}_2^v$ . If  $c_1 \in \mathcal{B}_2^v$ , then  $c_2$  is generated with  $\mathcal{B}_2^v$ . Moreover, as all the triangles of  $\text{GAD}(c_2, c_3)$  are in  $\mathcal{B}_2^v$  and  $c_2 \in \text{span}(\mathcal{B}_2^v)$  then  $c_3$  is spanned with  $\mathcal{B}_2^v$ . This construction with functions GAD forms a kind of clique of functions GAD, one clique per node. A minimum cycle basis of  $G_2$  is the union of the minimum cycle bases of its connected components. Hence, a minimum cycle basis of  $G_2$  contains exactly one square  $c_{(v,w)}$  for each node  $v \in V$ .

In  $G_3$ , we consider each node  $v \in V$  one by one. We apply function  $\text{GAD}(c_{(w_1,v)}, c_{(w_2,v)})$  for each pair of predecessors  $(w_1, w_2)$  of  $v$ . Thus, we cannot select two incoming arcs of a same node  $v$ . A minimum cycle basis of  $G_3$  contains exactly one square  $c_{(w,v)}$  for each node  $v \in V$ .

In  $G_1$ , we consider a cycle basis  $\mathcal{B} = \{d_1, d_2, \dots, d_p\}$  of the non-oriented multigraph underlying  $G$ . For  $1 \leq i \leq p$ , we consider  $d_i$  from the cycle basis and we apply function  $\text{GAD}(c_a, a \in d_i)$ . We denote  $\text{GAD}_i$  the set of cycles obtained by the application of the function  $\text{GAD}(c_a, a \in d_i)$ .

We will now prove that in any minimum cycle basis  $\mathcal{B}_1$  of  $G_1$ , a subset of arcs  $c$  contains a cycle in  $G$  if and only if the associated squares  $c_a$  for  $a \in c$  are not subset cycles of  $\mathcal{B}_1$ . Then, the properties of the minimum cycle basis of  $G_2$  and  $G_3$  are enough to show that the sectioned squares define a Hamiltonian path from  $s$  to  $t$  in  $G$  if and only if the maximum intersection is of size  $V - 1$ .

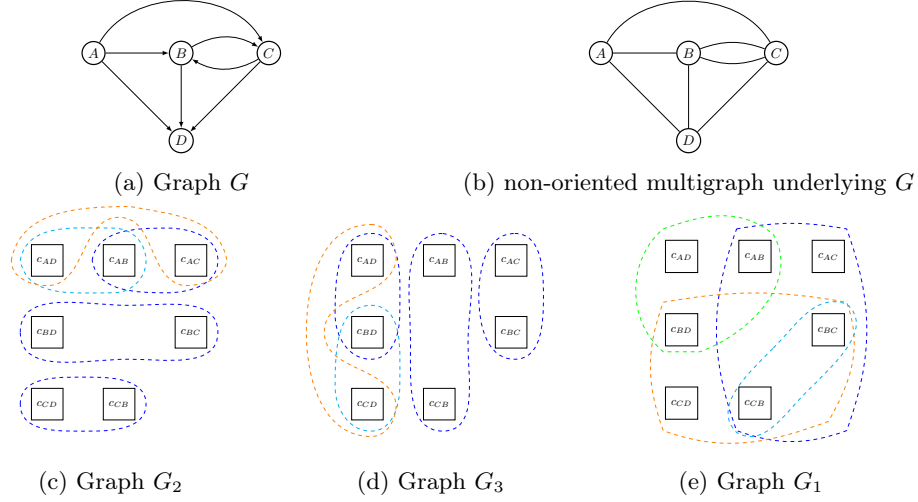


Fig. 2: Example of the construction of graphs  $G_1$ ,  $G_2$ , and  $G_3$  given an oriented graph  $G = (V, A)$ . The cycle basis considered to define  $G_1$  is the set of the inner faces in the representation of the underlying non-oriented multigraph drawn in the same plan as  $G$  in Figure 2a. In Figures 2c, 2d and 2e, the dashed lines represent the edges added between a set of cycles by a function GAD. Different colors are used to represent GAD around the same square for better readability.

1. If  $c$  is a cycle in  $G$  the associated squares  $c_a$  for  $a \in c$  are not in a minimum cycle basis of  $G_1$ . Considering  $c$  a cycle of  $G$ , let us prove that the set of triangles  $\mathcal{T}$  of  $G_1$  and the squares  $\mathcal{C} = \{c_a, a \in c\}$  are not linearly independent. Let  $\mathcal{B}_1 = \{e_1, e_2, \dots, e_l\}$  be a minimum cycle basis of  $G_1$ . We know  $c = \bigoplus_{i=1}^p \lambda_{\mathcal{B}}(d_i, c) \cdot d_i$  where  $\mathcal{B}$  refers to the basis of  $G$  used in the construction of  $G_1$ . Consider the following sum, where  $GAD_i$  denotes the set of cycles obtained by the application of the function  $GAD(c_a, a \in d_i)$ ,

$$S = \bigoplus_{i=1}^p \lambda_{\mathcal{B}}(d_i, c) \cdot \left( \bigoplus_{t \in GAD_i \cap \mathcal{T}} t + \bigoplus_{a \in GAD_i \cap \mathcal{C}} c_a \right) = \bigoplus_{i=1}^p \lambda_{\mathcal{B}}(d_i, c) \cdot \left( \bigoplus_{t \in GAD_i \cap \mathcal{T}} t + \bigoplus_{a \in d_i \cap \mathcal{C}} c_a \right)$$

Given that the faces of  $GAD_i$  correspond to the set of triangles of  $GAD_i \cap \mathcal{T}$  and  $\{c_a | a \in d_i\}$  then

$$\bigoplus_{t \in GAD_i \cap \mathcal{T}} t + \bigoplus_{a \in d_i \cap \mathcal{C}} c_a = \bigoplus_{a \in d_i \setminus c} c_a \text{ and, } S = \bigoplus_{i=1}^p \lambda_{\mathcal{B}}(d_i, c) \cdot \bigoplus_{a \in d_i \setminus c} c_a$$

Consider each arc of  $\bigcup_{i=1}^p d_i$ , if this arc is in  $c$  then it appears an odd number of times in the sum  $\bigoplus_{i=1}^p \lambda_{\mathcal{B}}(d_i, c) \cdot d_i$ ; otherwise it appears an even number of times. Thus, in the sum above, each arc  $a \notin d_i$ ,  $c_a$  appears an even number of times. Hence,  $S = 0$  and we can conclude that  $c \cup \mathcal{T}$  are not linearly independent.

2. If for a subset of arcs  $c$  the associated  $c_a$ , for  $a \in c$  are not subset cycles of  $\mathcal{B}_1$ , then  $c$  contains a cycle in  $G$ .

Given now a set of arcs  $c \subset A$  such that the set of triangles  $\mathcal{T}$  of  $G_1$  and  $\mathcal{C} = \{c_a, a \in c\}$  are not linearly independent. We assume  $c$  minimum, i.e., removing an arc from  $c$  makes the set we consider linearly independent. If  $|c| = 0$  then there is a contradiction because  $\mathcal{T}$  is linearly independent thus  $|c| > 0$ . We denote  $c = \{a_1, a_2, \dots, a_{|c|}\}$ . There exists a set  $\mathcal{T}' \subset \mathcal{T}$  such that  $\bigoplus_{t \in \mathcal{T}'} t + \bigoplus_{j=1}^{|c|} c_{a_j} = 0$ .

We now prove the fact that if  $\mathcal{T}'$  contains one triangle from  $GAD_i$  then  $\mathcal{T}'$  contains all the triangles of  $GAD_i$ . Otherwise, there is at least one edge  $e$  in  $GAD_i$  shared by two triangles  $t \in \mathcal{T}'$  and  $t' \notin \mathcal{T}'$ . Note that  $t$  and  $t'$  are the only cycles in  $\mathcal{T} \cup \mathcal{C}$  that contain  $e$ . Thus, the sum  $\bigoplus_{t \in \mathcal{T}'} t + \bigoplus_{j=1}^{|c|} c_{a_j}$  cannot equal zero because none of the cycles also cancel the edge  $e$  added by  $t$ . Without loss of generalities, we consider  $\mathcal{T}' = \bigcup_{i=1}^q GAD_i$  for some  $q \leq p$ . We can now show that  $\bigoplus_{i=1}^q d_i = c$ . For all  $i \leq q$ ,  $\bigoplus_{t \in \mathcal{T} \cap GAD_i} t + \bigoplus_{a \in d_i} c_a = 0$ . Thus,

$$\bigoplus_{t \in \mathcal{T}'} t + \bigoplus_{j=1}^{|c|} c_{a_j} + \bigoplus_{i=1}^q \left( \bigoplus_{t \in \mathcal{T} \cap GAD_i} t + \bigoplus_{a \in d_i} c_a \right) = 0$$

As every triangle of  $\mathcal{T}'$  is from  $GAD_{1 \leq i \leq q}$ , we get

$$\bigoplus_{j=1}^{|c|} c_{a_j} + \bigoplus_{i=1}^q \bigoplus_{a \in d_i} c_a = 0 \iff \bigoplus_{j=1}^{|c|} c_{a_j} = \bigoplus_{i=1}^q \bigoplus_{a \in d_i} c_a$$

As the squares of  $G_1$  are disjoint, this equality involves that  $\bigoplus_{i=1}^q d_i = c$  and means that the minimum cycle basis  $\mathcal{B}$  generates  $c$ . Hence,  $c$  is a cycle. Note that if  $c$  were not *minimum* as we have assumed it, then  $c$  would contain a cycle. Hence,  $c$  is a set of arcs containing a cycle of  $G$  if and only if  $\mathcal{T}$  and  $\mathcal{C} = \{c_a, a \in c\}$  are not linearly independent.

To conclude, any  $\mathcal{B}_1 \in \mathcal{MCB}(G_1)$  describes a covering tree of the associated non-oriented subgraph of  $G$ .

**Theorem 1.** *Problem Minimum Cycle Bases Intersection is NP-Complete for  $k \geq 3$ .*

*Proof.* The intersection of the minimum cycle bases of  $G_1$ ,  $G_2$  and  $G_3$  verifies the same properties as the independent of the intersection of the three matroids  $M_1$ ,  $M_2$  and  $M_3$ . In both cases, it is a union of elementary paths. Thus this intersection is of size  $|V| - 1$  if and only if  $G$  is Hamiltonian.

### 3 Exact and approximation algorithms for MCBI

This section proves that MCBI has a polynomial complexity for instances with at most two graphs. Then, this section provides a polynomial-time approximation algorithm  $\frac{1}{k}$  for MAX-MCBI where  $k$  is the number of graphs in the instance and proves that this ratio is the best feasible. We also provide results on the parameterized complexity of MCBI with respect to the size of the intersection.

For that, throughout this section, we make the assumption that we have two functions:

- CANDIDATESSETS( $G_1, G_2, \dots, G_k$ ): this function returns a polynomial-size set of cycles per given graph such that an optimal solution of MAX-MCBI can be extracted from these sets.
- MCBSET( $G, C$ ): this function returns a minimum cycle basis of  $G$  with the most cycles of  $C$ .

We propose a way to implement these functions in section 3.4.

#### 3.1 A polynomial algorithm when $k = 2$

In this section, the key idea is to prove that the problem MCBI can be reduced to the matroids intersection problem. In MCBI, we have two graphs,  $G_1$ , and  $G_2$ , and we look for two minimum cycle bases (one per graph) with the largest intersection. In the matroid intersection problem, we have two matroids  $(S, I_1)$  and  $(S, I_2)$  and we search for a maximum-size subset of  $S$  that belongs to  $I_1$  and  $I_2$ . In the case of the matroids, we know that this problem is polynomial in  $|S|$  assuming there exists a polynomial time algorithm to determine whether a subset of  $S$  belongs to  $I_1$  or  $I_2$  [7]. We can then reduce MCBI to this problem and then apply the polynomial algorithm to solve it. Two issues occur with this technique.



1. What is the ground set  $S$  in the corresponding matroids? We could use the cycles of  $G_1$  and  $G_2$  but the ground set of the matroids must be the same while the set of cycles of  $G_1$  and the set of cycles of  $G_2$  are different. Also, the number of cycles in a graph can be exponential, we could then not use an algorithm that is polynomial in  $|S|$ .
2. What is an independent in the corresponding matroids? We could set  $I_1$  and  $I_2$  as the minimum cycle bases of  $G_1$  and  $G_2$ , however then the first property of definition of a matroid is not verified because a subset of a cycle basis is not a cycle basis.

Let us first consider the second issue : given  $G$  a graph and  $S$  a subset of cycles of  $G$ ,  $I_G$  contains the subsets of  $S$  that belong to at least one minimum cycle basis of  $G$ . Hence,  $(S, I_G)$  defines a matroid that represents  $G$ .

**Lemma 5.** *Given  $f_1, f_2 \in I$  such that  $|f_1| < |f_2|$ , there exists  $e \in f_2 \setminus f_1$  such that  $f_1 \cup \{e\} \in I$ .*

*Proof.* Given  $f_1, f_2 \in I$  such that  $|f_1| < |f_2|$ , there exists two bases  $\mathcal{B}_1$  and  $\mathcal{B}_2$  where  $f_1 \subset \mathcal{B}_1$  and  $f_2 \subset \mathcal{B}_2$ . Let us show that there exists  $e \in f_2$  such that  $e \notin \text{SPAN}(f_1 \cup \{d \in \mathcal{B}_1, |d| < |e|\})$ . Thus, there would exist at least one cycle  $e$  of  $f_2$  and  $d \notin f_1$  such that  $|d| = |e|$  and  $\lambda_{\mathcal{B}_1}(d, e) = 1$ . Hence, according to Lemma 1 we can exchange  $d$  by  $e$  in  $\mathcal{B}_1$  to get a new cycle basis with the same weight containing  $f_1 \cup \{e\}$ . We will now prove that if all the  $e \in f_2$  are in  $\text{SPAN}(f_1 \cup \{d \in \mathcal{B}_1, |d| < |e|\})$  then we reach a conflict on the definition of  $\mathcal{B}_2$  as a cycle basis.

We have  $\forall e \in f_2, e = \bigoplus_{d \in \mathcal{B}_1 \setminus f_1, |d| < |e|} \lambda_{\mathcal{B}_1}(d, e) \cdot d + \bigoplus_{d \in f_1} \lambda_{\mathcal{B}_1}(d, e) \cdot d$

We denote  $X_e = \bigoplus_{d \in f_1} \lambda_{\mathcal{B}_1}(d, e) \cdot d$ . Hence,  $\forall e \in f_2, e = \bigoplus_{d \in \mathcal{B}_1 \setminus f_1, |d| < |e|} \lambda_{\mathcal{B}_1}(d, e) \cdot d + X_e$ .

Given that for all  $e \in f_2$ ,  $X_e \in \text{SPAN}(f_1)$  and  $|f_1| < |f_2|$  then the family  $\{X_e, e \in f_2\}$  is not linearly independent. Consider  $e_1, e_2, \dots, e_p$  such that  $\bigoplus_{i=1}^p X_{e_i} = 0$ , then :

$$\begin{aligned} e_1 + e_2 + \dots + e_p &= \bigoplus_{i=1}^p \bigoplus_{d \in \mathcal{B}_1 \setminus f_1, |d| < |e_i|} \lambda_{\mathcal{B}_1}(d, e) \cdot d + \bigoplus_{i=1}^p X_{e_i} \\ &= \bigoplus_{i=1}^p \bigoplus_{d \in \mathcal{B}_1 \setminus f_1, |d| < |e_i|} \lambda_{\mathcal{B}_1}(d, e) \cdot d \\ &= \bigoplus_{i=1}^p \bigoplus_{d \in \mathcal{B}_1 \setminus f_1, |d| < |e_i|} \lambda_{\mathcal{B}_1}(d, e) \cdot \left( \bigoplus_{f \in \mathcal{B}_2} (f, d) \cdot f \right) \end{aligned}$$

We assume that the  $e_i$  are sorted by weight and that  $|e_1| \leq |e_2| \leq \dots \leq |e_{q-1}| < |e_q| = |e_{q+1}| = \dots = |e_p|$ , then  $e_q + e_{q+1} + \dots + e_p = e_1 + e_2 + \dots + e_{q-1} + \bigoplus_{i=1}^p \bigoplus_{d \in \mathcal{B}_1 \setminus f_1, |d| < |e_i|} \lambda_{\mathcal{B}_1}(d, e) \cdot \left( \bigoplus_{f \in \mathcal{B}_2} (f, d) \cdot f \right)$ .

Both sums contain only cycles of  $\mathcal{B}_2$  however, the cycles in the right part have a weight strictly inferior to  $|e_p|$ . Indeed, if  $|d| < |e_i| \leq |e_p|$  and  $\lambda_{\mathcal{B}_2}(f, d) = 1$  then by Lemma 2,  $|f| \leq |d| < |e_p|$ . Hence, we have two sums of different cycles of  $\mathcal{B}_2$  equal, thus they are not linearly independent. This is in conflict with  $\mathcal{B}_2$  being a cycle basis. Hence, there exists  $e \in f_2$  such that  $e \notin \text{SPAN}(f_1 \cup \{d \in \mathcal{B}_1, |d| < |e|\})$ .

To answer now the first issue : we define  $S$  with the function  $\text{CANDIDATESSETS}(G_1, G_2)$ . This function returns two polynomial-size sets,  $S_1$  a subset of cycles of  $G_1$  and  $S_2$  a subset of cycles of  $G_2$ . Then,  $S$  is the intersection of  $S_1$  and  $S_2$ .

Thus we obtain two matroids  $M_1 = (S, I_{G_1})$  and  $M_2 = (S, I_{G_2})$  on which we can solve the matroid intersection problem. In order to use the algorithm of [7], we need a polynomial time algorithm to determine whether a subset of  $S$  is independent in  $M_1$  or  $M_2$ . To do so, we use the function  $\text{MCBSET}$ .

**Lemma 6.** *Function MCBSET can be used to determine if a subset of cycles is independent in the matroid  $(S, I_G)$  defined from a graph  $G$  in polynomial time with respect to the size of  $G$ .*

*Proof.* Function MCBSET takes a graph  $G = (V, E)$  and a set of cycles  $C$  and returns a minimum cycle basis of  $G$  with the most cycles from  $C$ . Given a matroid  $(S, I_G)$ ,  $f \in I_G$  if  $f \subset \text{MCBSET}(G, f)$ . The oracle is polynomial with respect to  $|V|$ .

**Theorem 2.** *Minimum Cycle Bases Intersection with  $k = 2$  is polynomial.*

*Proof.* The function  $\text{CANDIDATESSETS}(G_1, G_2)$  returns  $S_1$ , respectively  $S_2$ , a set of cycles of  $G_1$ , respectively  $G_2$ , such that a solution to MCBI can be extracted from it. Hence,  $S = S_1 \cap S_2$  contains a solution to MCBI, i.e., the largest subset of a minimum cycles basis shared in both  $G_1$  and  $G_2$ .

From a set  $S'$ , subset of a cycle basis we can easily extend it in two minimum cycle bases :  $\mathcal{B}_1 = \text{MCBSET}(G_1, S')$  and  $\mathcal{B}_2 = \text{MCBSET}(G_2, S')$ . Then, we have  $|\mathcal{B}_1 \cap \mathcal{B}_2| \geq |S'|$ .

Hence, solving the matroid intersection problem for  $M_1$  and  $M_2$ , enables us to find a solution to MCBI for  $G_1$  and  $G_2$ . According to Lemma 6 and the polynomial size of the set  $S$  induced by CANDIDATESSET, this algorithm is polynomial with respect to the size of  $G_1$  and  $G_2$ .

### 3.2 A polynomial $\frac{1}{k}$ -approximation algorithm

In this section we introduce Algorithm 1 an approximation for MAX-MCBI. The ratio of  $\frac{1}{k}$  obtained with Algorithm 1 is the best reachable for MAX-MCBI.

---

**Algorithm 1** Polynomial  $\frac{1}{k}$ -approximation algorithm for MAX-MCBI

---

```

1:  $\mathcal{B}_1, \mathcal{B}_2, \dots, \mathcal{B}_k \leftarrow \emptyset$ 
2:  $S_1, S_2, \dots, S_k \leftarrow \text{CANDIDATESSETS}(G_1, G_2, \dots, G_k)$ 
3: for each  $i$  from 1 to  $k$  do  $\mathcal{L}_i \leftarrow$  cycles of  $S_i$  ordered by weight.
4: for each  $l$  from 3 to  $n$  do
5:    $S'_1, S'_2, \dots, S'_k \leftarrow$  the cycles of  $\mathcal{L}_1, \mathcal{L}_2, \dots, \mathcal{L}_k$  of weight  $l$ 
6:   while  $\bigcap_{i=1}^k S'_i \neq \emptyset$  do add any cycle of  $\bigcap_{i=1}^k S'_i$  to all the set  $\mathcal{B}_i$  and then remove it from  $\mathcal{L}_i$  and  $S'_i$ , remove also the cycles generated by  $\mathcal{B}_i, \forall 1 \leq i \leq k$ .
7:   for each  $i$  from 1 to  $k$  do
8:     while  $S'_i \neq \emptyset$  do add any cycle of  $S'_i$  to  $\mathcal{B}_i$  and remove it from  $\mathcal{L}_i$  and  $S'_i$ , remove also the cycles now generated by  $\mathcal{B}_i$ 
9: return  $\mathcal{B}_1, \mathcal{B}_2, \dots, \mathcal{B}_k$ 
    
```

---

**Theorem 3.** *Problem MAX-MCBI is  $\frac{1}{k}$ -approximable.*

*Proof.* Algorithm 1 is polynomial as it consists of a loop running for a polynomial number of iterations; each iteration consists of elementary operations on sets or lists (searching for shortest paths in a graph, searching for linear combinations of cycles). Ordering at Line 3 is only done once and assure that Line 5 consists in elementary operations.

Lemma 8 states that there exists an optimal solution  $\mathcal{B}_1^*, \mathcal{B}_2^*, \dots, \mathcal{B}_k^*$  of MAX-MCBI such that  $\mathcal{B}_i^* \subseteq S_i$ . Let  $\mathcal{B}_1, \mathcal{B}_2, \dots, \mathcal{B}_k$  be the solution returned by Algorithm 1.

Let us prove that, for any  $l \in \mathbb{N}$ , the number of cycles of weight  $l$  in  $\bigcap_{i=1}^k \mathcal{B}_i^*$  is at most  $k$  times the number of such cycles in  $\bigcap_{i=1}^k \mathcal{B}_i$ . We denote the subset of cycles  $c \in \mathcal{B}_i$  with  $\omega(c) \leq l$  by  $\mathcal{B}_i(l)$ . As a consequence

of Lemma 3,  $\text{span}(\mathcal{B}_i(l)) = \text{span}(\mathcal{B}_i^*(l))$ . This gives us that  $\mathcal{S}'_i$ , defined at Line 5, verifies  $\mathcal{B}_i^*(l) \subseteq \mathcal{S}'_i$ . Then, starting the first iteration of the loop at Line 6, we have  $\bigcap_{i=1}^k \mathcal{B}_i^*(l) \subseteq \bigcap_{i=1}^k \mathcal{S}'_i$ .

Let  $p$  be the number of cycles in  $\bigcap_{i=1}^k \mathcal{B}_i^*(l)$ . Let us now show that at least  $\frac{p}{k}$  cycles are in  $\bigcap_{i=1}^k \mathcal{B}_i$ . As a consequence of Lemma 4 and because they span the same subset of cycles of  $G_i$ ,  $|\mathcal{B}_i(l)| = |\mathcal{B}_i^*(l)| = p$ . Let  $q$  be the number of cycles of length  $l$  in  $\mathcal{B}_i$ . We consider  $\Delta_i = c_1, \dots, c_q$  such that (i)  $\Delta_i \subseteq \mathcal{B}_i$ , (ii)  $|\Delta_i| = q$  and (iii)  $\forall 1 \leq j \leq k, c_j \in \Delta_i$  verifies  $\omega(c_j) = l$ . Thus, the set  $\{\mathcal{B}_i(l-1) \cup \Delta_i\}$  cannot generate more than  $|\mathcal{B}_i(l-1)| + |\Delta_i|$  cycles of length at most  $l$  in  $\mathcal{B}_i^*$ . As before,  $\text{span}(\mathcal{B}_i(l-1)) = \text{span}(\mathcal{B}_i^*(l-1))$  and thus at most  $q$  cycles have weight  $l$ .

To conclude, at the beginning of the  $(q+1)$ -th iteration of the loop at Line 6,  $\bigcap_{i=1}^k \mathcal{B}_i$  contains  $q$  cycles of length  $l$  and thus  $\mathcal{B}_i$  generates at most  $q$  cycles of length  $l$  of  $\mathcal{B}_i^*$ . Thus, we have  $|\bigcap_{i=1}^k \mathcal{S}'_i \cap \bigcap_{i=1}^k \mathcal{B}_i^*| \geq p - q \cdot k$ . Finally, if  $\bigcap_{i=1}^k \mathcal{S}'_i = \emptyset$ , then we have  $q \geq \frac{p}{k}$ , which means that at least  $\frac{p}{k}$  cycles are in  $\bigcap_{i=1}^k \mathcal{B}_i$  at the end of the loop.

We now show in Theorem 4 that there exists a reduction from the Maximum Independent Set problem (denoted here as MIS) to MCBI such that a solution of one problem is a solution to the other, namely, a strict reduction. Hence, as stated by the inapproximability result of MIS [20], it proves that the ratio of  $\frac{1}{k}$  is the best reachable for MAX-MCBI.

**Theorem 4.** *Unless  $P = NP$ , for every  $0 < \varepsilon$ , there is no polynomial approximation algorithm with ratio  $\frac{1}{k^{1-\varepsilon}}$ .*

*Proof.* Let us prove that there exists a strict reduction from MIS problem to MCBI where  $k = n$ . Let  $G = (V, E)$  be an instance of MIS with  $n = |V|$ . By the Vizing Theorem [16], using a polynomial greedy algorithm, edges of  $G$  can be covered with at most  $n$  disjoint matchings. The list of matchings can be completed with empty sets to get  $n$  matchings. Let  $M_1, M_2, \dots, M_n$  be those matchings. For each  $v \in V$ , a cycle  $c_v$  of length four is defined. Let us build an instance  $f(G) = (G_1, G_2, \dots, G_n)$  of MCBI as follows:  $\forall 1 \leq i \leq n$ , and each edge  $[v, w] \in M_i$ , we apply the procedure  $\text{GAD}(c_v, c_w)$  in  $G_i$  (described in Section 2.1).

Let  $\mathcal{B}_1, \mathcal{B}_2, \dots, \mathcal{B}_n$  be a feasible solution of  $f(G)$ . Let us set  $g(\mathcal{B}_1, \mathcal{B}_2, \dots, \mathcal{B}_n) = \{v | c_v \in \bigcap_{i=1}^n \mathcal{B}_i\}$ . Note that  $f$  and  $g$  can be computed in polynomial time. Thus as stated by Property 1 of the procedure GAD, any basis  $\mathcal{B}_i \in \mathcal{MCB}(G_i)$  contains all the cycles  $c_v$ , for  $v \in V$ , except one cycle  $c_v$  or  $c_w$  per edge  $[v, w] \in M_i$ .

The set  $\mathcal{S} = g(\mathcal{B}_1, \mathcal{B}_2, \dots, \mathcal{B}_n)$  is an independent set of  $G$  and  $|\mathcal{S}| = |\bigcap_{i=1}^n \mathcal{B}_i|$ . Given a maximum independent set  $\mathcal{S}^*$ ,  $\forall 1 \leq i \leq n$ , a cycle basis of  $G_i$  can be obtained by completing the set  $\{c_v | v \in \mathcal{S}^*\}$ , the size of the intersection of those bases is at least  $|\mathcal{S}^*|$ . Consequently, if  $\mathcal{B}_1^*, \mathcal{B}_2^*, \dots, \mathcal{B}_n^*$  is an optimal solution of  $f(G)$ ,  $|\mathcal{S}^*| \leq |\bigcap_{i=1}^n \mathcal{B}_i^*| = |g(\mathcal{B}_1^*, \mathcal{B}_2^*, \dots, \mathcal{B}_n^*)| \leq |\mathcal{S}^*|$ . The two optimal solutions have the same values. Thus, for every feasible solution  $\mathcal{B}_1, \mathcal{B}_2, \dots, \mathcal{B}_n$  of  $f(G)$ , we have  $|\mathcal{S}^*|/|\mathcal{S}| = |\bigcap_{i=1}^n \mathcal{B}_i^*|/|\bigcap_{i=1}^n \mathcal{B}_i|$ . Hence,  $(f, g)$  is a strict reduction.

### 3.3 Parameterized complexity of MCBI with respect to $\gamma$

As a reminder, given a set of  $k$  graphs  $G_{1 \leq i \leq k}$ , problem MCBI decides if there exists  $\mathcal{B}_i \in \mathcal{MCB}(G_i)$  such that  $|\bigcap_{i=1}^k \mathcal{B}_i| \geq \gamma$ .

**Theorem 5.** *MCBI is XP with respect to  $\gamma$ .*

*Proof.* Let  $\mathcal{S}_1, \mathcal{S}_2, \dots, \mathcal{S}_k$  be the sets returned by  $\text{CANDIDATESSETS}(G_1, G_2, \dots, G_k)$ . Note that each set contains a polynomial number of cycles.

We consider a function similar to the method proposed in Algorithm 1. This function, denoted  $\text{MCBSET}$ , takes a graph  $G$  and a cycle set  $\mathcal{C}$  as arguments. It returns a minimum cycle basis of  $G$  that contains the most cycles of  $\mathcal{C}$ . In Section 3.4, we propose a method to implement it.

Let us present the following algorithm. Start by enumerating all tuples of  $\gamma$  cycles in  $\mathcal{S}_i$ . Then to verify if there exists a tuple  $t$  included in the intersection of the bases, proceed as follows: for each tuple  $t$  appearing in all the sets  $\mathcal{S}_i; \forall 1 \leq j \leq k, \mathcal{B}_j^t = \text{MCBSET}(G_j, t)$ . Then, check for each tuple  $t$ , if there exists  $\mathcal{B}_j^t \in \text{MCB}(G_j)$  such that  $t \subset \mathcal{B}_j^t$ . If and only if such bases exist in every graph,  $G_1, G_2, \dots, G_k$  is a YES-instance.

By Lemma 8, an optimal solution is contained in  $\mathcal{S}_1, \mathcal{S}_2, \dots, \mathcal{S}_k$ , hence, our algorithm is correct. The exponential part of the algorithm is the enumeration of all tuples of each set  $\mathcal{S}_i$  in time  $|\mathcal{S}_i|^\gamma$ . Thus the complexity is XP with respect to  $\gamma$ .

**Theorem 6.** *MCBI is W[1]-Hard with respect to  $\gamma$ .*

*Proof.* The reduction in the proof of Theorem 4 is also an FPT reduction. This time, from the problem of searching an Independent Set of size  $K'$  in an undirected graph to MCBI parameterized with  $\gamma = K'$ . The result follows from the W[1]-hardness of the first problem [6].

### 3.4 Polynomial algorithms for MCBSET and CANDIDATESSET

This section proposes two implementable algorithms for the functions MCBSET and CANDIDATESSETS that we assumed in Section 3. These two algorithms are built from the polynomial algorithm of Horton [12].

**Maximizing intersection with a given set** We propose a method to implement function MCBSET in Algorithm 2. Given a graph  $G$  and a set of cycles  $\mathcal{C}$ , the function returns  $\mathcal{B} \in \text{MCB}(G)$  such that  $\mathcal{B}$  contains the most cycles of  $\mathcal{C}$ .

---

**Algorithm 2** Maximizing minimum cycle basis intersection with a set

---

```

1: function MCBSET( $G = (V, E), \mathcal{C}$ ) :  $\mathcal{B}$ 
2:   compute  $\mathcal{B}'$  a minimum cycle basis of  $G$ 
3:    $\mathcal{B} \leftarrow \emptyset, \mathcal{S} \leftarrow \mathcal{C} \cup \mathcal{B}'$ 
4:    $\mathcal{L} \leftarrow$  cycles  $c \in \mathcal{S}$  sorted by weight, in case of equality  $c \in \mathcal{C}$  is ordered first
5:    $size \leftarrow |E| - |V| + x$  where  $x$  is the number of connected component in  $G$ 
6:   while  $|\mathcal{B}| < size$  do
7:     consider  $c$  the first element of  $\mathcal{L}$  and remove it from  $\mathcal{L}$ 
8:     if  $c$  is linearly independent to  $\mathcal{B}$  then
9:       add  $c$  to  $\mathcal{B}$ 
10:  return  $\mathcal{B}$ 
    
```

---

In the Algorithm 2 we considered that  $\mathcal{C}$  was a subset of cycles of  $G$ . Otherwise, we will first have to reduce the set  $\mathcal{C}$  to the cycles of  $G$ . The particular sort applied in Algorithm 2 at Line 4 is referred to as SAH (sort of amended Horton).

**Lemma 7.** *Finding a minimum cycle basis that maximizes its intersection with a given set of cycles is done in polynomial-time with Algorithm 2.*

*Proof.* Algorithm 2 is polynomial as it consists of consecutive polynomial operations and an elementary loop running for a polynomial number of iterations. The minimum cycle basis  $\mathcal{B}'$  at Line 2 can be computed in polynomial time with [12,1].

Let  $\mathcal{C}$  be a given set of cycles, and  $\mathcal{B}$  be the result of Algorithm 2. Let  $\mathcal{B}^* = (e_1, e_2, \dots, e_{\mu(G)})$  be, among the minimum cycle bases of  $G$  maximizing  $|\mathcal{B}^* \cap \mathcal{C}|$ , a basis maximizing  $|\mathcal{B} \cap \mathcal{B}^*|$ . Let us prove that

$|\mathcal{B} \cap \mathcal{C}| = |\mathcal{B}^* \cap \mathcal{C}|$ . For that, we consider that  $\mathcal{B}^*$  is ordered by SAH. Let us suppose that  $|\mathcal{B} \cap \mathcal{C}| < |\mathcal{B}^* \cap \mathcal{C}|$  and let  $e_i \in \mathcal{B}^*$  be the first cycle of  $\mathcal{B}^* \cap \mathcal{C} \setminus \mathcal{B}$ .

$\mathcal{B}$  is a minimum cycle basis as it is the  $\mu(G)$  independent smallest cycles of  $\mathcal{S}$ . Thus, there exists  $\Delta \subseteq \mathcal{B}$  such that  $e_i = \bigoplus_{d \in \Delta} \lambda_{\mathcal{B}}(d, e_i) \cdot d$ .

We know  $\Delta \not\subseteq \mathcal{B}^*$  because  $e_i \in \mathcal{B}^*$  and otherwise  $\mathcal{B}^*$  cannot be a cycle basis. As stated in Lemma 2,  $\forall d \in \Delta$ , we have  $\omega(d) \leq \omega(e_i)$ . Thus, if  $\forall d \in \Delta$ ,  $\omega(d) < \omega(e_i)$  then  $\mathcal{B}^*$  cannot be in  $\mathcal{MCB}(G)$  because  $e_i$  could be exchanged in  $\mathcal{B}^*$  by a cycle  $d \in \Delta$ . According to Lemma 1, the basis such obtained has a lower weight than  $\mathcal{B}^*$ . Hence, it must exist  $\hat{d} \in \Delta$  with  $\omega(\hat{d}) = \omega(e_i)$ .

Now, if  $\omega(\hat{d}) = \omega(e_i)$ , as  $\hat{d} \in \mathcal{B}$  and  $e_i \notin \mathcal{B}$  then  $\hat{d}$  has been ordered before  $e_i$  by SAH. Also, given that  $e_i \in \mathcal{C}$  then  $\hat{d} \in \mathcal{C}$ ; If  $e_i$  is exchanged in  $\mathcal{B}^*$  with  $\hat{d}$  then the basis such obtained has : (i) the same weight as  $\mathcal{B}^*$ , (ii) the same intersection size with  $\mathcal{C}$ , (iii) but a larger intersection with  $\mathcal{B}$ . Such a basis cannot exist by definition of  $\mathcal{B}^*$ , hence by contradiction we have our results.

As a conclusion, Algorithm 2 provides an optimal solution in polynomial-time.

**Enumeration of cycles** We present Algorithm 3 that given a set of graphs  $G_{1 \leq i \leq k}$ , returns  $k$  sets of cycles, one set  $\mathcal{S}_i$  per graph  $G_i$ . Lemma 8 proves that an optimal solution of MAX-MCBI is a subset of these sets. For  $1 \leq i \leq k$ , we call  $\mathcal{S}_i$  a set of *candidates cycles*.

---

**Algorithm 3** Enumerate the sets of candidates cycles containing an optimal solution of MAX-MCBI

---

```

1: function CANDIDATESSETS( $G_1, G_2, \dots, G_k$ )
2:    $\mathcal{S}_1, \mathcal{S}_2, \dots, \mathcal{S}_k \leftarrow \emptyset$ 
3:    $G = (V, E) \leftarrow (V, \bigcap_{i=1}^k E_i)$  with  $G_i = (V, E_i)$ 
4:   for  $u \in V$ ,  $[v, w] \in E$  do add to all the sets  $\mathcal{S}_i$  the cycle consisting in  $[v, w]$ , a shortest path from  $u$  to  $v$  and a shortest path from  $u$  to  $w$  in  $G$ .
5:   for  $[u, v] \in E$ ,  $[w, x] \in E$  do add to all the sets  $\mathcal{S}_i$  two cycles consisting in  $[u, v]$ ,  $[w, x]$ , one with the shortest paths from  $u$  to  $w$  and from  $v$  to  $x$  in  $G$ , and a second with the shortest paths from  $u$  to  $x$  and from  $v$  to  $w$  in  $G$ .
6:   for  $i$  from 1 to  $k$  do compute  $\mathcal{B}_i \in \mathcal{MCB}(G_i)$  and add all the cycles of  $\mathcal{B}_i$  to  $\mathcal{S}_i$ .
7:   return  $\mathcal{S}_1, \mathcal{S}_2, \dots, \mathcal{S}_k$ 
8:   ▷ At Lines 4 and 5, a cycle is added only if the shortest paths are disjoint.

```

---

**Lemma 8.** *Given an instance  $\{G_1, \dots, G_k\}$  of MAX-MCBI and  $\mathcal{S}_1, \dots, \mathcal{S}_k$  the sets returned by CANDIDATESSETS( $G_1, \dots, G_k$ ), there exists an optimal solution  $\{\mathcal{B}_1^*, \dots, \mathcal{B}_k^*\}$  such that for all  $1 \leq i \leq k$ ,  $\mathcal{B}_i^* \subseteq \mathcal{S}_i$ .*

*Proof.* Let  $\mathcal{B}_1^*, \mathcal{B}_2^*, \dots, \mathcal{B}_k^*$  be any optimal solution to MAX-MCBI for  $G_{1 \leq i \leq k}$ . Let us build a solution  $\mathcal{B}_1, \dots, \mathcal{B}_k$  such that  $\mathcal{B}_i \subseteq \mathcal{S}_i$  as follows. For that, we suppose that  $\bigcap_{i=1}^k \mathcal{B}_i^* \subseteq \bigcap_{i=1}^k \mathcal{S}_i$ . Let  $\mathcal{B}_i \in \mathcal{MCB}(G_i)$  be by applying MCBSET( $G_i, \bigcap_{i=1}^k \mathcal{B}_i^*$ ) with the only change that the intermediary basis  $\mathcal{B}'$  computed at Line 2 (Algorithm 2) must be a subset of  $\mathcal{S}_i$ . Thus,  $\forall 1 \leq j \leq k$ , we have  $\bigcap_{i=1}^k \mathcal{B}_i^* \subseteq \mathcal{B}_j$ . Finally, the such defined  $\mathcal{B}_1, \dots, \mathcal{B}_k$  is solution of MAX-MCBI as it verifies  $\mathcal{B}_i \subseteq \mathcal{S}_i$  and  $\bigcap_{i=1}^k \mathcal{B}_i = \bigcap_{i=1}^k \mathcal{B}_i^*$ .

Now to verify that such a solution is achievable, let us prove that  $\bigcap_{i=1}^k \mathcal{B}_i^* \subseteq \bigcap_{i=1}^k \mathcal{S}_i$ . For that, we suppose that there exists a cycle  $c$  such that  $c \in \bigcap_{i=1}^k \mathcal{B}_i^*$  and  $c \notin \bigcap_{i=1}^k \mathcal{S}_i$ . For any such cycle  $c$ , there exist 3 cycles (as illustrates in Figure 3)  $c_1, c', c_2$  with  $\omega(c_1) < \omega(c)$ ,  $\omega(c') = \omega(c)$  and  $\omega(c_2) < \omega(c)$  such that  $c_1 + c' + c_2 = c$  and  $c' \in \bigcap_{i=1}^k \mathcal{S}_i$ . Any such cycle  $c$  can be replaced by  $c'$  in all the cycle bases such that  $\mathcal{B}_i^* \cup \{c'\} \setminus \{c\} \in \mathcal{MCB}(G_i)$ .

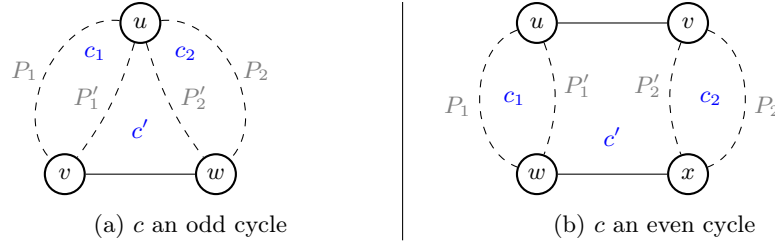


Fig. 3: Example of three cycles such that  $c_1 + c' + c_2 = c$  and  $\omega(c') = \omega(c)$ . If  $|P_1| < |P_2|$  then there exists a node  $x$  on  $P_2$ , such that the cycle passing by  $x, v$  and  $w$  matches to one of these examples.

Now, let us prove that such a  $c'$  exists. Consider that  $c$  is an odd cycle, then  $c$  was not considered in the loop at Line 4 of Algorithm 3. Thus, there exist shortest paths  $P_1, P_2, P_1'$  and  $P_2'$  such as shown on Figure 3 with  $|P_1| = |P_1'| = |P_2| = |P_2'|$ . As  $c$  belongs to all minimum cycle bases then  $\omega(c') = \omega(c)$ , otherwise they could not be minimum (consequence of Lemma 2). Then, such a cycle  $c'$  exists and has been added to  $\bigcap_{i=1}^k \mathcal{S}_i$  by Algorithm 3 instead of  $c$ . Similarly, if  $c$  is an even cycle, the same result is obtained by using the loop at Line 5 in Algorithm 3 that build the cycle  $c'$ . Note that if  $P_1$  is equal to  $P_1'$  then  $c_1$  is empty, and respectively, if  $P_2$  equals  $P_2'$  then  $c_2$  is empty.

To conclude, if such a cycle  $c$  exists then, the cycle  $c'$  exists. Thus, new minimum cycle bases solution of MAX-MCBI can be built by replacing  $c$  with  $c'$  in all  $\mathcal{B}_i^*$  with  $1 \leq i \leq k$  (this operation is possible by Lemma 1); the resulting bases have the same intersection size. Repeat this operation until  $\bigcap_{i=1}^k \mathcal{B}_i^* \subseteq \bigcap_{i=1}^k \mathcal{S}_i$ .

## 4 Concluding remarks

This paper introduces the problem of searching for close minimum cycle bases between graphs with the same vertices and study its complexity and approximability.

Despite the hard complexity of MCBI, the polynomial case when  $k = 2$ , the approximation algorithm and the function MCBSET give us the hope that we can solve this problem with relative efficiency in the specific case of the molecular dynamics trajectories discussed in the introduction. Indeed, in these sequences, between a graph and its successor, few edges appear or disappear [5]. The use of local search could be efficient in computation time and performance in the case of sets of graphs from a same dynamic.

Also, our proof on subsets of minimum cycle bases that can be independent in matroid might be generalized to the minimum cycle basis of matroids [10]. Hence, we may generalize some of our results to this case.

## References

1. Amaldi, E., Iuliano, C., Jurkiewicz, T., Mehlhorn, K., Rizzi, R.: Breaking the  $o(m^2n)$  barrier for minimum cycle bases. In: Algorithms - ESA 2009. pp. 301–312. European Symposium on Algorithms, Springer Berlin Heidelberg (2009)
2. Berge, C.: Graphs and hypergraphs. North-Holland Pub. Co. (1973)
3. Berger, F., Gritzmann, P., de Vries, S.: Minimum cycle bases and their applications. In: Algorithmics of Large and Complex Networks, pp. 34–49. Springer Berlin Heidelberg (2009). [https://doi.org/10.1007/978-3-642-02094-0\\_2](https://doi.org/10.1007/978-3-642-02094-0_2)
4. Blumenthal, D.B., Bougleux, S., Gamper, J., Brun, L.: Ring based approximation of graph edit distance. In: Bai, X., Hancock, E.R., Ho, T.K., Wilson, R.C., Biggio, B., Robles-Kelly, A. (eds.) Structural, Syntactic, and Statistical Pattern Recognition. pp. 293–303. Springer International Publishing, Cham (2018)

5. Bougueroua, S., Spezia, R., Pezzotti, S., Vial, S., Quessette, F., Barth, D., Gaigeot, M.P.: Graph theory for automatic structural recognition in molecular dynamics simulations. *The Journal of Chemical Physics* **149**(18), 184102 (nov 2018). <https://doi.org/10.1063/1.5045818>
6. Downey, R., Fellows, M.: Parameterized complexity. Springer-Verlag New York (1999). <https://doi.org/10.1007/978-1-4612-0515-9>
7. Edmonds, J.: Matroid intersection. In: Hammer, P., Johnson, E., Korte, B. (eds.) *Discrete Optimization I*, *Annals of Discrete Mathematics*, vol. 4, pp. 39–49. Elsevier (1979). [https://doi.org/https://doi.org/10.1016/S0167-5060\(08\)70817-3](https://doi.org/https://doi.org/10.1016/S0167-5060(08)70817-3), <https://www.sciencedirect.com/science/article/pii/S0167506008708173>
8. Gaüzère, B., Brun, L., Villemin, D.: Relevant cycle hypergraph representation for molecules. In: Kropatsch, W.G., Artner, N.M., Haxhimusa, Y., Jiang, X. (eds.) *Graph-Based Representations in Pattern Recognition*. pp. 111–120. Springer Berlin Heidelberg, Berlin, Heidelberg (2013)
9. Gleiss, P.M., Stadler, P.F., Wagner, A., Fell, D.A.: Relevant cycles in chemical reaction networks. *Advances in complex systems* **4**(02n03), 207–226 (jun 2001). <https://doi.org/10.1142/s0219525901000140>
10. Golynski, A., Horton, J.D.: A polynomial time algorithm to find the minimum cycle basis of a regular matroid. In: Penttonen, M., Schmidt, E.M. (eds.) *Algorithm Theory — SWAT 2002*. pp. 200–209. Springer Berlin Heidelberg, Berlin, Heidelberg (2002)
11. Guignon, V., Chauve, C., Hamel, S.: An edit distance between rna stem-loops. In: Consens, M., Navarro, G. (eds.) *String Processing and Information Retrieval*. pp. 335–347. Springer Berlin Heidelberg, Berlin, Heidelberg (2005)
12. Horton, J.D.: A polynomial-time algorithm to find the shortest cycle basis of a graph. *SIAM Journal on Computing* **16**(2), 358–366 (apr 1987)
13. Ilemo, S.N., Barth, D., David, O., Quessette, F., Weisser, M.A., Watel, D.: Improving graphs of cycles approach to structural similarity of molecules. *PLOS ONE* **14**(12), e0226680 (dec 2019). <https://doi.org/10.1371/journal.pone.0226680>
14. Karp, R.M.: *Reducibility among Combinatorial Problems*, pp. 85–103. Springer US (1972)
15. Kavitha, T., Liebchen, C., Mehlhorn, K., Michail, D., Rizzi, R., Ueckerdt, T., Zweig, K.A.: Cycle bases in graphs characterization, algorithms, complexity, and applications. *Computer Science Review* **3**(4), 199–243 (nov 2009). <https://doi.org/10.1016/j.cosrev.2009.08.001>
16. Misra, J., Gries, D.: A constructive proof of vizing's theorem. *Information Processing Letters* **41**(3), 131–133 (mar 1992). [https://doi.org/10.1016/0020-0190\(92\)90041-s](https://doi.org/10.1016/0020-0190(92)90041-s)
17. Vismara, P.: Union of all the minimum cycle bases of a graph. *The Electronic Journal of Combinatorics* **4**(1) (jan 1997). <https://doi.org/10.37236/1294>
18. Welsh, D.J.: *Matroid theory*. Courier Corporation (2010)
19. Yang, L., Liu, Y., Hu, X., Wang, P., Li, X., Wu, J.: Graph-based analysis of rna secondary structure similarity comparison. *Complexity* **2021**, 1–15 (2021). <https://doi.org/10.1155/2021/8841822>
20. Zuckerman, D.: Linear degree extractors and the inapproximability of max clique and chromatic number. In: *Proceedings of the Thirty-Eighth Annual ACM Symposium on Theory of Computing*. p. 681–690. STOC '06, Association for Computing Machinery, New York, NY, USA (2006). <https://doi.org/10.1145/1132516.1132612>, <https://doi.org/10.1145/1132516.1132612>