



HAL
open science

Complexity and algorithms for Swap median and relation to other consensus problems

Luís Cunha, Thiago Lopes, Arnaud Mary

► **To cite this version:**

Luís Cunha, Thiago Lopes, Arnaud Mary. Complexity and algorithms for Swap median and relation to other consensus problems. 2024. hal-04886139

HAL Id: hal-04886139

<https://hal.science/hal-04886139v1>

Preprint submitted on 14 Jan 2025


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.



Distributed under a Creative Commons Attribution 4.0 International License

Complexity and algorithms for SWAP MEDIAN and relation to other consensus problems

Luís Cunha¹ ✉ 🏠 

Fluminense Federal University, Niterói, Brazil

Thiago Lopes ✉

Fluminense Federal University, Niterói, Brazil

Arnaud Mary ✉ 🏠

Laboratoire de Biométrie et Biologie Évolutive, Université Claud Bernard Lyon 1, France

Abstract

Genome rearrangements are events in which large blocks of DNA exchange pieces during evolution. The analysis of such events is a tool for understanding evolutionary genomics, based on finding the minimum number of rearrangements to transform one genome into another. In a general scenario, more than two genomes are considered and we have new challenges. The MEDIAN problem consists in finding, given three permutations and a distance metric, a permutation s that minimizes the sum of the distances between s and each input. We study the MEDIAN problem over *swap* distances in permutations, for which the computational complexity has been open for almost 20 years (Eriksen, *Theor. Compt. Sci.*, 2007). We consider this problem through some branches. We associate median solutions and interval convex sets, where the concept of graph convexity inspires the following investigation: Does a median permutation belong to every shortest path between one of the pairs of input permutations? We are able to partially answer this question, and as a by-product we solve a long open problem by proving that the SWAP MEDIAN problem is NP-hard. Furthermore, using a similar approach, we show that the CLOSEST problem, which seeks to minimize the maximum distance between the solution and the input permutations, is NP-hard even considering three input permutations. This gives a sharp dichotomy into the P vs. NP-hard approaches, since considering two input permutations the problem is easily solvable and considering any number of input permutations it is known to be NP-hard since 2007 (Popov, *Theor. Compt. Sci.*, 2007). In addition, we show that SWAP MEDIAN and SWAP CLOSEST are APX-hard problems.

2012 ACM Subject Classification CCS, Mathematics of computing, Discrete mathematics

Keywords and phrases swap median, swap closest, permutations, complexity.

Digital Object Identifier [10.4230/LIPIcs.conf.2025](https://doi.org/10.4230/LIPIcs.conf.2025).

Funding *Luís Cunha*: FAPERJ grant JCNE E-26/201.372/2022, CNPq Universal 406173/2021-4

1 Introduction

Ancestral reconstruction is a classic problem in comparative genomics. This problem is based on consensus word analysis with broad applicability [6, 10, 14]. Genome rearrangement problems study large-scale mutations on a set of DNAs in living organisms, and have been extensively studied in computational biology and computer science fields for decades. From a mathematical point of view, a genome can be represented as strings or permutations. Watterson et al. [22] proposed genome rearrangement problems interpreted as transforming one permutation into another by a minimum number of operations depending on the possible allowed rearrangements, i.e. the chosen metric. This model makes the following assumptions: the order of genes in each genome is known; all genomes we compare share the same gene set;

¹ corresponding author



all genomes contain a single copy of each gene; and all genomes consist of a single chromosome. Thus, genomes can be modeled as permutations once each gene encodes an integer.

Permutations are sequences of distinct integers, and a *swap operation* is an exchange between two elements on the permutation. Given two permutations with the same number of elements, the *swap distance* is the smallest number of swaps that transform one permutation into the other. The swap distance problem is solved in polynomial time, given by the function of the number of the algebraic cycles of the permutation [8]. Many other metrics related to genome rearrangements have been studied and have a variety of applications in computational biology [5, 7, 9, 18, 20].

For each of these metrics, it is possible to make a correlation between sets of elements, such as the CLOSEST GENOME problem. This problem consists in finding a genome that minimizes the maximum distance with respect to the input set based on the chosen metric [5, 17]. It is well known that the CLOSEST STRING problem over Hamming distance is NP-hard even for binary strings. When dealing with permutations, Cunha et al. [5] settled NP-hardness results on CLOSEST PERMUTATION regarding breakpoints and block-interchanges. Furthermore, several FPT (*fixed parameterized tractable*) results were obtained regarding swap and other metrics for the CLOSEST PERMUTATION problems [4].

GRAPH CONVEXITY is a field in which different combinatorial problems can be defined according to the application [13]. In general, we are dealing with sets of vertices that are to be increased step by step, following some prescribed rule, until the process stabilizes in a *convex set* (the “closure”). Cunha and Protti [7] proposed a connection between graph convexity and genome rearrangements. This concept is viewed as permutations that can be generated from a given input set, and we relate it to the MEDIAN problem, presented next.

Given a set of three genomes and a metric, the goal of the MEDIAN problem is to find a solution genome that minimizes the sum of the distances between the solution and all elements that belong to the input set. MEDIAN solution of a given set is another relevant way to summarize a set of genomes, which affects many approaches to finding ancestral genomes [1, 3, 4, 5, 7, 12, 15]. Our study focuses on the SWAP MEDIAN problem over permutations, where the computational complexity was open for almost 20 years [8]. We present the following contributions:

- We settle that the SWAP MEDIAN problem is NP-hard, which solves a long open problem [8].
- We prove that the SWAP CLOSEST problem is NP-hard even considering three input permutations, which is a stronger result compared to the well known NP-hardness where the number of input permutations is arbitrary [17].
- We prove that SWAP MEDIAN and SWAP CLOSEST are APX-hard problems.

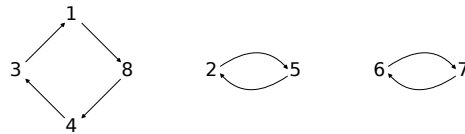
This article is organized as follows: In Section 2, we present preliminaries concerning SWAP MEDIAN, GRAPH CONVEXITY and we discuss the relation between MEDIAN and CONVEXITY problems; In Section 3 we show that SWAP MEDIAN is NP-hard, by showing that it is a difficult task even to decide whether a solution reaches the lower bound of a given input, and as a by-product we prove that SWAP CLOSEST is NP-hard even when considering three input permutations; In Section 4, we prove that both problems are APX-hard; Finally, in Section 5 we present a conclusion for further investigation.

2 Preliminaries

An *alphabet* Σ is a non empty set of letters, and a *string* over Σ is a finite sequence of letter of Σ . A *permutation* of size n is a particular string with a single occurrence of each element

from 1 to n , considering that the permutation is a bijection of the set $\{1, 2, \dots, n\}$ with itself. Hence, we have the permutation $\pi_n = [\pi(1) \pi(2) \dots \pi(n)]$. A permutation π_n can usually be denoted as π , when the size is implicitly known. Given a metric M and two permutations π and σ of same size, $d_M(\pi, \sigma)$ is the distance between π and σ . Since we can relabel σ to transform it to the *identity permutation* $\iota_n = [1, 2, \dots, n]$ by applying $\sigma\sigma^{-1} = \iota$, we have $d_M(\pi, \sigma) = d_M(\pi\sigma^{-1}, \iota)$. Therefore, we compute the distance between two permutations where one of them is ι , denoting $d_M(\gamma, \iota)$, or simply by $d_M(\gamma)$.

Permutations can also be represented by each element followed by its image. For example, given a set $\{1, 2, 3\}$, the sequence (1 2 3) maps 1 into 2, 2 into 3, and 3 into 1, corresponding to the permutation [2 3 1]. This representation is not unique; for instance, (2 3 1) and (3 1 2) are equivalent. Permutations are composed of one or more algebraic cycles, where each *algebraic cycle* of a permutation π is a representation of a domain i , followed by its image $\pi(i)$, followed by getting the image of $\pi(i)$ as the next element, i.e., $\pi(\pi(i))$, and so on. We continue this process until we reach a repeated element. This procedure uniquely defines the permutation. We denote by $c(\pi)$ the number of algebraic cycles of π and a *k-cycle* is the number of elements of an algebraic cycles of π . For example, given $\pi = [8 \ 5 \ 1 \ 3 \ 2 \ 7 \ 6 \ 4] = (1 \ 8 \ 4 \ 3)(2 \ 5)(6 \ 7)$, we have $c(\pi) = 3$, with two 2-cycles and one 4-cycle. Those cycles can be described as the *algebraic cycles graph* $G(\pi)$, as described in Figure 1.



■ **Figure 1** Algebraic cycles graph $G(\pi)$ of the permutation $\pi = [8 \ 5 \ 1 \ 3 \ 2 \ 7 \ 6 \ 4]$.

A swap involving elements a and b such that a and b are in the same cycle is a swap that *breaks* the cycle into two, whereas if a and b belong to different cycles, the swap of these elements *unites* the two cycles [10]. Thus, the swap distance of a permutation π is determined as follows: $d_{\text{swap}}(\pi) = n - c(\pi)$, where $c(\pi)$ is the number of algebraic cycles of π .

MEDIAN problem by a metric M

The decision version of the MEDIAN problem by the metric M is formalized as follows:

MEDIAN BY THE METRIC M (M-MEDIAN)

Instance: Set of permutations $\{\pi_1, \pi_2, \pi_3\}$, and a non negative integer f .

Question: Is there a permutation σ such that $\sum_{i=1,2,3} d_M(\pi_i, \sigma) \leq f$?

Note that, similar to distance problems, for the MEDIAN problem we can consider π_1 as $\iota = [1 \ 2 \dots n]$ to make the study easier, and relabel the other two permutations accordingly. For example, if π_1 is [3 1 2], π_2 is [2 1 3], and π_3 is [1 3 2], we can relabel π_1 as [1 2 3], π_2 as [3 2 1], and π_3 as [2 1 3]. Moreover, as a consequence of the triangular inequality, every median f for permutations π_1, π_2, π_3 satisfies $f \geq \frac{d_M(\pi_1, \pi_2) + d_M(\pi_1, \pi_3) + d_M(\pi_2, \pi_3)}{2}$. A permutation σ that achieves the median is called the *median solution* (or *median permutation*).

There are several studies on MEDIAN problems with different focuses. Silva et al. [20] treated this problem under a restricted rearrangement metric called C_4 -distance and developed heuristics and linear programming formulations. Shao and Moret [19] considered the double-cut-and-join metric, and Tannier et al. [21] showed a number of algorithms for

multichromosomal genomes. Eriksen [8] developed properties concerning solutions on the REVERSAL MEDIAN, for which it is known to be NP-complete [3], and also on the SWAP MEDIAN, for which in that paper it was called TRANSPOSITION MEDIAN, but with the same meaning as swaps, and its computational complexity has been unknown until now.

Convexity problems

A finite convexity space is a pair (V, \mathcal{C}) consisting of a finite set V and a family \mathcal{C} of subsets of V such that $\emptyset \in \mathcal{C}$, $V \in \mathcal{C}$, and \mathcal{C} is closed under intersections. Members of \mathcal{C} are called *convex sets*. Let \mathcal{P} be a collection of paths of a graph G , and let $I_{\mathcal{P}} : 2^{V(G)} \rightarrow 2^{V(G)}$ be a function (called *interval function*) such that: $I_{\mathcal{P}}(S) = S \cup \{z \notin S \mid \exists u, v \in S \text{ such that } z \text{ lies in an } uv\text{-path } P \in \mathcal{P}\}$.

Different choices of \mathcal{P} lead to interval functions with quite different behavior. Such functions are naturally associated with special convexity spaces (the so-called PATH CONVEXITIES). For example, if \mathcal{P} contains exactly all shortest paths in a graph, then the corresponding interval function is naturally associated with the well-known GEODESIC CONVEXITY; if \mathcal{P} is the collection of induced paths, then the corresponding interval function is associated with the MONOPHONIC CONVEXITY; and there are many other examples in the literature. The STEINER CONVEXITY is the collection of vertices of *Steiner trees* with three permutations as input, along the vertices of the shortest paths between the inputs. The *Steiner tree* with three input permutations consists in finding a permutation (a vertex in the corresponding graph) that minimizes the total weight of its edges [2]. In other words, a median solution of a given set of three permutations belongs to a solution of the STEINER problem [3]. Based on this, the STEINER CONVEXITY gives the collection of all elements that form possible solutions for the Steiner tree given a metric.

As a consequence of the just presented definitions, if the GEODESIC interval convex set $(I_G(S))$ is equal to the MONOPHONIC interval convex set $(I_M(S))$, then they are equal to the STEINER interval convex set $(I_S(S))$. Overall, $I_G(S) \subseteq I_S(S) \subseteq I_M(S)$ [2].

We present two questions below as a first attempt to investigate a median solution.

First question: Does a median permutation always belong to one of the shortest paths between at least two of the input permutations?

Second question: Can a brute-force approach find a median permutation in feasible time if the first question is true?

Using the algebraic cycles, we analyze how many sequences there are in which any operation is performed on two elements of the same cycle. This is important to focus on, because in all the cases we analyzed, the median solutions belong to the shortest paths of at least one pair of the inputs. So if we could prove that the number of permutations in optimal paths is a polynomial function, then it might be a nice approach to analyze the first question. Proposition 1 shows the opposite with respect to the second question, even if the first is true.

► **Proposition 1.** *With respect to an instance of the swap median problem, there is an exponential number of permutations that belong to all shortest paths.*

Proof. Let ι , π and σ be the input permutations of size n . For each pair of such permutations there is a corresponding algebraic cycle, one from ι to π , one from π to σ , and one from σ to ι . Notice that the algebraic cycle from ι to π consists only of 2 cycles. So there are $\frac{n}{2}$ 2-cycles. Since the swaps applied in any optimal path break cycles, there are $(\frac{n}{2})!$ sequences of swaps, i.e. optimal paths, to transform π into ι .

Note that in each of these sequences there is at least one permutation that is not contained in any other optimal path. In general, there are $\frac{n}{2}$ operations that can be applied in such

a way that all cycles become 1-cycles, where $(\frac{n}{2})!$ is the number of possible choices of $\frac{n}{2}$ 2-cycles. Each new sequence adds at least one distinct permutation, otherwise two distinct sequences would have all permutations the same, so these two permutations would not be distinct. So there is an exponential number of permutations belonging to the shortest paths from ι to π , which is sufficient to guarantee an exponential number of permutations also from ι to σ and from π to σ . This concludes the proof. ◀

To answer the first question just presented, we partition all instances into three different types. Let π_1 , π_2 , and π_3 be the inputs and f be the median solution: Recall that, as a consequence of the triangular inequality, $f \geq \frac{d_M(\pi_1, \pi_2) + d_M(\pi_1, \pi_3) + d_M(\pi_2, \pi_3)}{2}$. Then, compared to such a lower bound, we have the following possible solutions:

1. Type 1) f is equal to the lower bound. In this case, the median permutation belongs to the shortest path between the inputs.

If f does not match the lower bound, there are two options:

1. Type 2) It may be possible that the solution permutation is on the optimal path of at least one and at most two pairs between π_1 and π_2 , π_1 and π_3 , and π_2 and π_3 .
2. Type 3) It may be possible that the solution permutation is not on the optimal path between π_1 and π_2 , between π_1 and π_3 nor between π_2 and π_3 . Thus, the permutation solution will be at least 3 units above the lower bound.

Hence, to give a negative answer for the first question, we need to find an instance that can be classified as type 3 according to these types.

3 SWAP MEDIAN and SWAP CLOSEST are NP-complete even considering 3 input permutations

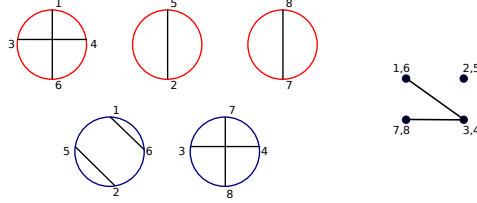
The NP-completeness proof of SWAP MEDIAN is obtained by dealing with a simpler question, as follows. Let π_1, π_2 and π_3 be a given input. Hence: Does there exist a solution σ for which the sum of the distances between σ and all other input permutations is equal to $\frac{d_{\text{swap}}(\pi_1, \pi_2) + d_{\text{swap}}(\pi_1, \pi_3) + d_{\text{swap}}(\pi_2, \pi_3)}{2}$?

If such a permutation exists, then σ belongs to a shortest path between the three pairs of input permutations. Therefore, let x and y be integers such that $d_{\text{swap}}(\pi_1, \sigma) = \beta$, $d_{\text{swap}}(\pi_2, \sigma) = x$, $d_{\text{swap}}(\pi_2, \pi_3) = x + y$. By solving this system, the question above is the same as to decide if there exists β such that $\beta = \frac{d_{\text{swap}}(\pi_1, \pi_2) + d_{\text{swap}}(\pi_1, \pi_3) - d_{\text{swap}}(\pi_2, \pi_3)}{2}$.

Assume without loss of generality that $\pi_1 = \iota$. Let \mathcal{M} be the set of optimal swaps sequence to transform π_1 into the median σ . We consider the two cycles decomposition $c(\pi_1, \pi_2)$ as the *red circles* and $c(\pi_1, \pi_3)$ as the *blue circles*. To achieve the lower bound for the median, a swap applied over π_1 must break cycles with respect to π_2 and to π_3 at once. Hence, a *possible swap* is a swap for which both endpoints are in a same cycle for both decompositions. Moreover, if σ reaches the lower bound, and if u, v and x, y are two swaps in a shortest swaps sequence from π_1 to σ that belong to the same cycle of either $c(\pi_1, \pi_3)$ or $c(\pi_1, \pi_2)$, then u, v and x, y should not cross.

Thus, there exists a median solution σ that reaches the lower bound, if and only if there exists a sequence \mathcal{M} of possible swaps from π_1 to σ of size $|\mathcal{M}| = \beta = \frac{d_{\text{swap}}(\pi_1, \pi_2) + d_{\text{swap}}(\pi_1, \pi_3) - d_{\text{swap}}(\pi_2, \pi_3)}{2}$ such that both endpoints of each swap of \mathcal{M} are both in a same red circle and in a same blue circle and such that every pair of swaps of \mathcal{M} contained in a same circle do not cross in that circle.

Hence, those β possible swaps form an independent set of size β in the graph we call *2-circles-intersection* graph, defined as follows: the vertices are possible swaps, i.e. each arc where the endpoints are in both red and blue circles; two vertices are adjacent if their corresponding arcs cross in any circle of the blue or the red model (see Figure 2 for an example).



■ **Figure 2** In the left: Two circle decomposition, one in red and another in blue. Each arc is a possible swap it can be applied to break a cycle. In the right: 2-circles-intersection graph, where each arc (possible swap) is a vertex and two crossing arcs are adjacent in the graph.

Let $Subd_2(G)$ be the graph obtained from a graph G where for each edge $vu \in E(G)$, add two vertices x_{vu} and x_{uv} , and replace the edge vu to a path v, x_{vu}, x_{uv}, u . The $Subd_2(G)$ graph is also referred as a *2 subdivision graph*.

► **Theorem 2.** *The 2-circles-intersection graph class contains the 2 subdivision graph class.*

Proof. Given a 2 subdivision graph, we formulate it as the circle decomposition, described in Construction 1, below.

► **Construction 1.** *Given a graph G and its 2 subdivision graph $Subd_2(G)$, where for each edge vu of G there is a path v, vu, uv, u in $Subd_2(G)$, obtain two circular models as follows (also illustrated in Figure 3). The first model is given as:*

1. For each vertex $v \in V(G)$, define a circular model C_v^1 with the arc s_v, e_v ;
2. For each vertex vu adjacent to v in $Subd_2(G)$, add an arc s_{vu}, e_{vu} intersecting only s_v, e_v in C_v^1 . This means that if there are more than one neighbor of v , the corresponding arcs will be parallel between them, but all of them will intersect s_v, e_v in C_v .

Thus, the first model C^1 is defined as the union of C_v^1 , for each $v \in V(G)$. These circles are called as the red circles.

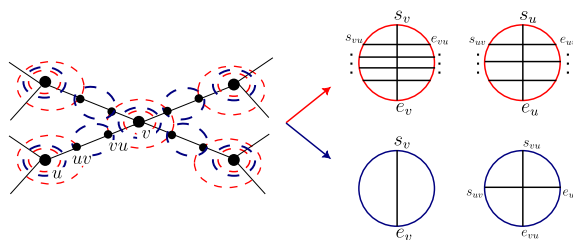
The second model is given as:

1. For each vertex $v \in V(G)$, define a circular model C_v^2 with the arc s_v, e_v ;
2. For each edge $vu \in G$, define a distinct model C_{vu}^2 with two intersecting arcs, which are s_{vu}, e_{vu} and s_{uv}, e_{uv} .

Thus, the second model C^2 is defined as the union of C_v^2 , for each $v \in V(G)$ and the union of C_{vu}^2 for each edge $vu \in E(G)$. These circles are called as the blue circles.

Since any 2 subdivision graph $Subd_2(G)$ can be seen as two circles decomposition, hence $Subd_2(G)$ can be formulated as a 2-circles-intersection graph H , by setting the set of arcs of the model to correspond to the set of vertices of H and two arcs that cross in the model correspond to adjacent vertices in H . ◀

Since the 2-circles-intersection graphs contain the 2 subdivisions graphs (Theorem 2), and MAXIMUM INDEPENDENT SET problem (MIS) is known to be NP-hard even for 2 subdivisions graphs [16], next we prove that determining a median solution that achieves the lower bound is equivalent to find an independent set in those graphs.



■ **Figure 3** In the left: A 2-subdivision of a graph with red and blue dashed ellipses corresponding to the two circular models we obtain the two algebraic cycles graphs. In the right: Parts of two circular models. In top are the red circles C_v^1 and C_u^1 of C^1 , and in bottom are the blue circles C_v^2 and $C_{v,u}^2$ of C^2 .

► **Construction 2.** Let $Subd_2(G)$ be the 2-subdivision of a graph G with its two circular models C^1 and C^2 , as described in the proof of [Theorem 2 \(Construction 1\)](#). We obtain an instance of the median problem π_1, π_2, π_3 , where each permutation has size $2|V(Subd_2(G))|$, and:

- $\pi_1 = \iota$;
- The algebraic cycles graph of π_2 (of π_3) with respect to π_1 is C^1 (is C^2). Hence, an algebraic cycle of π_2 (of π_3) is formed by the elements reading following the clockwise direction in a cycle of C^1 (of C^2).

Based on [Construction 2](#), we have that each arc of a circle model indeed corresponds to a possible swap to be applied over the associated permutation in order to approach to the median solution.

► **Theorem 3.** Let G be a graph with n vertices and m edges. There is an independent set of size α in $Subd_2(G)$ if and only if there is a sequence of $\alpha = \frac{3m+k}{2}$ swaps, for $1 \leq k \leq 2n+4m$, in the resulted instance of [Construction 2](#) from π_1 to σ such that σ achieves the lower bound for the median problem.

Proof. Assume π_1, π_2, π_3 are permutations from [Construction 2](#) where they come from a 2-subdivision graph $Subd_2(G)$, where $|V(G)| = n$ and $|E(G)| = m$. Since each permutation has size $2|V(Subd_2(G))|$, it implies that such a size is $2(n+2m)$. Hence, in order to decide the existence of a permutation σ that reaches the lower bound for the median problem, we must check the existence of α such that $\alpha = \frac{d_{\text{swap}}(\pi_1, \pi_2) + d_{\text{swap}}(\pi_1, \pi_2) - d_{\text{swap}}(\pi_2, \pi_3)}{2} = \frac{2n+4m - |C^1| + 2n+4m - |C^2| - (2n+4m - |c(C^1, C^2)|)}{2}$, where $c(C^1, C^2)$ is the algebraic cycles decompositions between π_2 and π_3 . Since for each vertex in G we created a red circle in [Construction 1](#), we have that $|C^1| = n$. Similarly, for each vertex and for each edge in G we created a blue circle in [Construction 1](#), then we have that $|C^2| = n+m$. This implies that $\alpha = \frac{3m + |c(C^1, C^2)|}{2}$, where $c(C^1, C^2)$ can be obtained by starting at an element u of cycle at C^1 , take the element v that has an directed edge to u in C^1 , i.e. the element in the backwards position, and then take the element w that v has an edge from v in a cycle at C^2 . Hence, the cycle will be described as u followed by w , and the process continue until all elements have been settled. Since the number of algebraic cycles is between 1 and the size of the permutation, which is $2n+4m$, hence we conclude the proof. ◀

It has been shown in [16] that deciding whether the 2-subdivision of a graph G has an independent set of size k is NP-complete. We show here that it remains true even for large k .

► **Theorem 4.** Given a graph G with n vertices and m edges and $k \geq \frac{3m}{2}$, it is NP-complete to decide whether $Subd_2(G)$ has an independent larger than k .

Proof. It has been shown in [16] that $\alpha(\text{Subd}_2(G)) = \alpha(G) + m$ and so that the Maximum independent set problem is NP-hard in 2-subdivision graphs. Observe that the maximum independent set problem is also NP-hard on graphs having more than $2n$ edges.

Given $k \geq \frac{3m}{2}$ and given a graph G we claim that G has an independent set of size k' if and only if $\text{Subd}_2(G \cup I_{k-k'-m})$ has an independent set of size k where $G \cup I_{k-k'-m}$ is the graph obtained from G by adding $k-k'-m$ isolated vertices. Observe first since we can assume that $k' \leq n \leq \frac{m}{2}$, we have $k-k'-m \geq \frac{3m}{2} - \frac{m}{2} - m \geq 0$ and so the graph $G \cup I_{k-k'-m}$ is well defined. Now we have $\alpha(\text{Subd}_2(G \cup I_{k-k'-m})) = \alpha(G \cup I_{k-k'-m}) + m = \alpha(G) + k - k' - m + m = \alpha(G) + k - k'$. So $\alpha(G) \geq k'$ if and only if $\alpha(\text{Subd}_2(G \cup I_{k-k'-m})) \geq k' + k - k' \geq k$. ◀

► **Corollary 5.** *Given three permutations π_1, π_2, π_3 it is NP-complete if there is a permutation σ that achieves the lower bound α for the swap median problem, where $\alpha = \frac{d_{\text{swap}}(\pi_1, \pi_2) + d_{\text{swap}}(\pi_1, \pi_3) - d_{\text{swap}}(\pi_2, \pi_3)}{2}$.*

Reducing MEDIAN to CLOSEST

The polynomial reduction presented in Corollary 5 allows us to show that not only the SWAP CLOSEST problem is NP-hard for three input permutations, but also that the CLOSEST problem with a constant number of input permutations is NP-hard, as obtained next.

► **Definition 6.** *Given π^1 with p elements and π^2 with q elements, the union of π^1 and π^2 is a permutation $\pi^1 \uplus \pi^2$ with $p+q$ elements such that $\pi^1 \uplus \pi^2 = [\pi_1^1, \pi_2^1, \dots, \pi_p^1, (\pi_1^2 + p), (\pi_2^2 + p), \dots, (\pi_q^2 + p)]$. Permutations π^1 and π^2 are called parts of the union.*

► **Lemma 7.** *Given permutations π^1 and π^2 , we have that $d_{\text{swap}}(\pi^{1,2}) = d_{\text{swap}}(\pi^1) + d_{\text{swap}}(\pi^2)$.*

Proof. Assuming that π^1 has p elements and π^2 has q elements, since $p+1$ is greater than all elements of π^1 , the algebraic cycles of $\pi^1 \uplus \pi^2$ is obtained by the union of the cycles of π^1 and of π^2 and shifting each element of π^2 by p . Therefore, we have $d_{\text{swap}}(\pi^{1,2}) = p+q - c(\pi^1) - c(\pi^2) = d_{\text{swap}}(\pi^1) + d_{\text{swap}}(\pi^2)$. ◀

Now, we develop a polynomial transformation from a general instance π^1, π^2, π^3 to a special one $\pi''^1, \pi''^2, \pi''^3$ in order to have all of these three pairs with a same distance. Based on this, we are able to guarantee that a closest solution of the new instance $\pi''^1, \pi''^2, \pi''^3$ reaches the lower bound if and only if there is a sequence of swaps from π^1 that reaches this same lower bound. Furthermore, since it is NP-complete to decide if there is a sequence of swaps from π^1 that approaches to π^2 and π^3 to achieve to lower bound (Corollary 5), we conclude from Corollary 12 that the SWAP CLOSEST problem is NP-complete even if there are three input permutations.

► **Construction 3.** *Let π^1, π^2, π^3 be permutations of size n where $d_{\text{swap}}(\pi^1, \pi^2) = \beta$, $d_{\text{swap}}(\pi^1, \pi^3) = \delta$, and $d_{\text{swap}}(\pi^2, \pi^3) = \alpha$. We transform π^1, π^2, π^3 into $\pi''^1, \pi''^2, \pi''^3$, resp. by the following unions:*

$$\begin{aligned} \pi''^1 &= \pi^1 \uplus \iota_{\beta-\alpha+1} && \uplus [2, 3, 4, \dots, \delta - \alpha + 1, 1]; \\ \pi''^2 &= \pi^2 \uplus \iota_{\beta-\alpha+1} && \uplus \iota_{\delta-\alpha+1}; \\ \pi''^3 &= \pi^3 \uplus [2, 3, 4, \dots, \beta - \alpha + 1, 1] && \uplus [2, 3, 4, \dots, \delta - \alpha + 1, 1]. \end{aligned}$$

► **Lemma 8.** *Let π^1, π^2, π^3 be permutations of size n where $d_{\text{swap}}(\pi^1, \pi^2) = \beta$, $d_{\text{swap}}(\pi^1, \pi^3) = \delta$ and $d_{\text{swap}}(\pi^2, \pi^3) = \alpha$. The permutations $\pi''^1, \pi''^2, \pi''^3$ obtained from Construction 3 satisfy $d_{\text{swap}}(\pi''^1, \pi''^2) = d_{\text{swap}}(\pi''^1, \pi''^3) = d_{\text{swap}}(\pi''^2, \pi''^3) = \beta + \delta - \alpha$.*

Proof. The algebraic cycle of the permutation $[2, 3, 4, \dots, n, 1]$ is $(1, 2, 3, \dots, n)$, which implies the swap distance of such a permutation equal to $n - 1$. From [Lemma 7](#), the distance between each pairwise between $\pi''^1, \pi''^2, \pi''^3$ can be obtained by summing the distances between each part separately of the unions. Therefore: $d_{\text{swap}}(\pi''^1, \pi''^2) = d_{\text{swap}}(\pi^1, \pi^2) + d_{\text{swap}}(\iota_{\beta-\alpha+1}) + d_{\text{swap}}([2, 3, 4, \dots, \delta - \alpha + 1, 1]) = \beta + 0 + \delta - \alpha + 1 - 1 = \beta + \delta - \alpha$; $d_{\text{swap}}(\pi''^1, \pi''^3) = d_{\text{swap}}(\pi^1, \pi^3) + d_{\text{swap}}([2, 3, 4, \dots, \beta - \alpha + 1, 1]) + d_{\text{swap}}([2, 3, 4, \dots, \delta - \alpha + 1, 1], [2, 3, 4, \dots, \delta - \alpha + 1, 1]) = \delta + \beta - \alpha + 1 - 1 + 0 = \beta + \delta - \alpha$; and $d_{\text{swap}}(\pi''^2, \pi''^3) = d_{\text{swap}}(\pi^2, \pi^3) + d_{\text{swap}}([2, 3, 4, \dots, \beta - \alpha + 1, 1]) + d([2, 3, 4, \dots, \delta - \alpha + 1, 1]) = \alpha + \beta - \alpha + 1 - 1 + \delta - \alpha + 1 - 1 = \beta + \delta - \alpha$. ◀

► **Lemma 9.** *Let $\pi''^1, \pi''^2, \pi''^3$ be permutations obtained from [Construction 3](#) where each pairwise has distance equal to c . Thus, there exists a sequence of $\frac{c}{2}$ swaps from π''^1 where each swap breaks cycle with respect to π''^2 and to π''^3 if and only if the closest solution of $\pi''^1, \pi''^2, \pi''^3$ is equal to $\frac{c}{2}$.*

Proof. Assume that there exists a closest solution of value $\frac{c}{2}$, which achieves the lower bound for that problem, because the lower bound is $\frac{\max_{i < j} d_{\text{swap}}(\pi''^i, \pi''^j)}{2}$ by the triangular inequality. Since the closest solution σ gives an upper bound on the distance it and each pair of input, therefore the median solution of $\pi''^1, \pi''^2, \pi''^3$ is at most $\frac{3c}{2}$. Moreover, a median solution is at least $\frac{3c}{2}$, because this bound is given by $\frac{\sum_{i < j} d_{\text{swap}}(\pi''^i, \pi''^j)}{2}$, then we have achieved the median solution to be equal to $\frac{3c}{2}$. Hence, since this solution achieves the lower bound, there is a sequence of independent swaps, which is a sequence of swaps that each of them breaks cycles from π''^1 to the solution in an optimum path to π''^2 and to π''^3 of size $c - \frac{c}{2} = \frac{c}{2}$.

Assume that there exists a sequence of $\frac{c}{2}$ swaps from π''^1 where each swap breaks cycle with respect to π''^2 and to π''^3 . Suppose that σ is the permutation obtained after those swaps. Now, we prove that $\max_{i \leq 3} d_{\text{swap}}(\pi''^i, \sigma) \leq \frac{c}{2}$. Since each one of those swaps approaches π''^1 to π''^2 and to π''^3 , we have that $d_{\text{swap}}(\sigma, \pi''^i) = d_{\text{swap}}(\pi''^i, \pi''^j) - \frac{c}{2} = \frac{c}{2}$. Hence, it achieves the lower bound of the closest problem, and so, we have that the closest solution is equal to $\frac{c}{2}$, as we desired. ◀

As a consequence of [Construction 3](#), any swap involving elements after the n -th position will join cycles in either π''^2 or π''^3 . Thus it proves [Lemma 10](#) below.

► **Lemma 10.** *Let π^1, π^2, π^3 be permutations of size n and $\pi''^1, \pi''^2, \pi''^3$ be permutations obtained by [Construction 3](#). Any sequence of swaps that approach π''^1 to σ , where σ is a closest solution, only contains swaps between elements from the position 1 to the position n .*

► **Corollary 11.** *The closest solution of $\pi''^1, \pi''^2, \pi''^3$ is equal to $\frac{c}{2}$, where c is the distance between any pairwise permutation, if and only if there is a sequence of swaps that approach π''^1 to π''^2 and π''^3 of size $\frac{c}{2}$.*

► **Corollary 12.** *It is NP-complete to decide if an input instance of the SWAP CLOSEST problem with three input permutations is equal to the lower bound of $\frac{\max_{i < j} d_{\text{swap}}(\pi^i, \pi^j)}{2}$.*

4 APX-hardness

Now, we deal with the problem to decide if it is possible to approximate a solution for the closest problem to be by a function of $1 + \epsilon$ of the lower bound, for any $\epsilon > 0$, and we prove that it is NP-complete. Hence, it implies that deciding if an instance has a solution closest solution that achieves the lower bound is APX-hard, as well for the SWAP MEDIAN, as consequence of the proof of [Theorem 13](#).

APPROXIMATING SWAP CLOSEST SOLUTION

Instance: Set of permutations $\{\pi_1, \pi_2, \pi_3\}$.

Question: Is there permutation σ whose value achieves the lower bound of $\frac{\max_{i < j} d_{\text{swap}}(\pi^i, \pi^j)}{2}$ or assert that it is smaller than $\frac{\max_{i < j} d_{\text{swap}}(\pi^i, \pi^j)}{2}(1 + \epsilon)$, for any $\epsilon > 0$?

► **Theorem 13.** APPROXIMATING SWAP CLOSEST SOLUTION is NP-complete.

Proof. Consider an instance π^1, π^2, π^3 whose the distance between each pairwise is the same value. Moreover, let us assume it is possible to approximate the closest solution to be on the ratio $1 + \epsilon$ of the lower bound, for which the value is OPT. Hence, it implies that from π^1 there is a sequence of $\text{OPT}(1 + \epsilon)$ swaps at approximates to π^2 and π^3 . Since the closest solution gives a maximum distance between a solution and each input permutation, we have that the median solution can be upper bounded by: $\sum_{i < j} d_{\text{swap}}(\pi^i, \pi^j) \leq 3 \max_i d_{\text{swap}}(\pi^i, \pi^j) \leq 3\text{OPT}(1 + \epsilon)2$, because by the triangular inequality $d_{\text{swap}}(\pi^i, \pi^j) \leq d_{\text{swap}}(\pi^i, \sigma) + d_{\text{swap}}(\sigma, \pi^j) = \text{OPT}(1 + \epsilon)2$. Thus, it implies that $\sum_{i < j} d_{\text{swap}}(\pi^i, \pi^j) \leq 6\text{OPT}(1 + \epsilon)$ and $\text{OPT} \geq \frac{\sum_{i < j} d_{\text{swap}}(\pi^i, \pi^j)}{6(1 + \epsilon)} = \frac{\sum_{i < j} d_{\text{swap}}(\pi^i, \pi^j)}{2} \cdot \frac{1}{(3 + \epsilon)}$. Now, considering the SWAP MEDIAN problem, in order to decide if there is a permutation that achieves the lower bound, we have that such a value is equal to $\beta = \frac{\sum_{i < j} d_{\text{swap}}(\pi^i, \pi^j) - 2d_{\text{swap}}(\pi^2, \pi^3)}{2} = \frac{\sum_{i < j} d_{\text{swap}}(\pi^i, \pi^j)}{2} - d_{\text{swap}}(\pi^2, \pi^3)$. Hence, in order to know when $\text{OPT} = \beta$, it occurs when $d_{\text{swap}}(\pi^2, \pi^3) = \frac{\sum_{i < j} d_{\text{swap}}(\pi^i, \pi^j)}{3}$, which is exactly the case we are considering, once we have assumed $d_{\text{swap}}(\pi^1, \pi^2) = d_{\text{swap}}(\pi^1, \pi^3) = d_{\text{swap}}(\pi^2, \pi^3)$. Thus, since the lower bound for the median β is achieved when a sequence of β swaps approach π^1 to π^2 and to π^3 per each operation, it also occurs for the SWAP CLOSEST problem when it achieves its lower bound. Hence it implies that there is a sequence of $\text{OPT}(1 + \epsilon)$ swaps from π^1 to π^2 and to π^3 where each of such operation breaks cycle on the two cycles decomposition ($c(\pi^1, \pi^2)$ and $c(\pi^1, \pi^3)$). Hence, those swaps form an independent set on the graph structure defined in Construction 2 obtained from the instance π^1, π^2, π^3 . As proved in Theorem 2, the 2-subdivision graph class is contained on the graph class we defined in Construction 2. Moreover, it is well known that the MAXIMUM INDEPENDENT SET problem is APX-hard on 2-subdivisions graphs [11]. Therefore, the $\text{OPT}(1 + \epsilon)$ swaps would imply a PTAS for MAXIMUM INDEPENDENT SET on those graphs, which is a contradiction. ◀

► **Corollary 14.** There is no $(1 + \epsilon)$ -approximation for the SWAP CLOSEST problem and for the SWAP MEDIAN problem, for any $\epsilon > 0$ even considering three input permutations.

5 Conclusion

This paper solves a long open problem by proving that the SWAP MEDIAN problem is NP-hard even considering three input permutations. In addition, we show that the SWAP CLOSEST problem is NP-hard also considering three input permutations, which is a stronger result than those already known. We also show that both problems are APX-hard. For further work, we intend to prove (or find a counterexample) that for any three input permutations, there exists a median solution that belongs to a shortest path of at least one input pair, which implies that a median solution always belongs (or does not belong) to the geodesic Steiner convexity.

References

- 1 Martin Bader. The transposition median problem is NP-complete. *Theor. Comput. Sci.*, 412(12-14):1099–1110, 2011.
- 2 José Cáceres, Alberto Márquez, and María Luz Puertas. Steiner distance and convexity in graphs. *European Journal of Combinatorics*, 29(3):726–736, 2008.
- 3 Alberto Caprara. The reversal median problem. *INFORMS J. Comput.*, 15(1):93–113, 2003.
- 4 Luís Cunha, Ignasi Sau, and Uéverton Souza. On the Complexity of the Median and Closest Permutation Problems. In *24th International Workshop on Algorithms in Bioinformatics (WABI 2024)*, volume 312 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 2:1–2:23, 2024.
- 5 Luís Felipe I Cunha, Pedro Feijão, Vinícius F dos Santos, Luis Antonio B Kowada, and Celina MH de Figueiredo. On the computational complexity of closest genome problems. *Discret. Appl. Math.*, 274:26–34, 2020.
- 6 Luís Felipe I Cunha, Luis Antonio B Kowada, Rodrigo de A. Hausen, and Celina MH de Figueiredo. Advancing the transposition distance and diameter through lonely permutations. *SIAM J. Discret. Math.*, 27(4):1682–1709, 2013.
- 7 Luís Felipe I Cunha and Fábio Protti. Genome rearrangements on multigenomic models: Applications of graph convexity problems. *J. Comput. Biol.*, 26(11):1214–1222, 2019.
- 8 Niklas Eriksen. Reversal and transposition medians. *Theoretical Computer Science*, 374(1-3):111–126, 2007.
- 9 Pedro Feijao and Joao Meidanis. Scj: a breakpoint-like distance that simplifies several rearrangement problems. *IEEE/ACM Trans. Comput. Biol. Bioinform.*, 8(5):1318–1329, 2011.
- 10 Guillaume Fertin, Anthony Labarre, Irena Rusu, Stéphane Vialette, and Eric Tannier. *Combinatorics of genome rearrangements*. MIT press, 2009.
- 11 Mathew C Francis, Daniel Gonçalves, and Pascal Ochem. The maximum clique problem in multiple interval graphs. *Algorithmica*, 71:812–836, 2015.
- 12 Maryam Haghghi and David Sankoff. Medians seek the corners, and other conjectures. In *BMC bioinformatics*, volume 13, pages 1–7. Springer, 2012.
- 13 L. D. Penso, F. Protti, D. Rautenbach, and U. Souza. Complexity analysis of P_3 -convexity problems on bounded-degree and planar graphs. *Theor. Comput. Sci.*, 607:83–95, 2015.
- 14 Pavel Pevzner. *Computational Molecular Biology: An Algorithmic Approach*. MIT press, 2000.
- 15 Itsik Pe’er and Ron Shamir. The median problems for breakpoints are NP-complete. In *Elec. Colloq. on Comput. Complexity*, volume 71, 1998.
- 16 Svatopluk Poljak. A note on stable sets and colorings of graphs. *Commentat. Math. Univ. Carol.*, 15:307–309, 1974.
- 17 V Yu Popov. Multiple genome rearrangement by swaps and by element duplications. *Theor. Comput. Sci.*, 385(1-3):115–126, 2007.
- 18 David Sankoff. Genome rearrangement with gene families. *Bioinformatics*, 15(11):909–917, 1999.
- 19 Mingfu Shao and Bernard ME Moret. On the DCJ median problem. In *Combinatorial Pattern Matching: 25th Annual Symposium, CPM 2014, Moscow, Russia, June 16-18, 2014. Proceedings 25*, pages 273–282. Springer, 2014.
- 20 Helmuth Silva, Diego Rubert, Eloi Araujo, Eckhard Steffen, Daniel Doerr, and Fábio Martinez. Algorithms for the genome median under a restricted measure of rearrangement. *RAIRO-Oper. Res.*, 57(3):1045–1058, 2023.
- 21 Eric Tannier, Chunfang Zheng, and David Sankoff. Multichromosomal median and halving problems under different genomic distances. *BMC Bioinformatics*, 10(1):1–15, 2009.
- 22 Geoffrey A Watterson, Warren J Ewens, Thomas Eric Hall, and Alexander Morgan. The chromosome inversion problem. *J. Theor. Biol.*, 99(1):1–7, 1982.