

Algorithms for the Maximum Eulerian Cycle Decomposition Problem

Pedro Olímpio Pinheiro

Institute of Computing, University of Campinas
Campinas, Brazil
pedro.pinheiro@students.ic.unicamp.br

Alexsandro Oliveira Alexandrino

Institute of Computing, University of Campinas
Campinas, Brazil
alexsandro@ic.unicamp.br

Andre Rodrigues Oliveira

Institute of Computing, University of Campinas
Campinas, Brazil
andrero@ic.unicamp.br

Cid Carvalho de Souza

Institute of Computing, University of Campinas
Campinas, Brazil
cid@ic.unicamp.br

Zanoni Dias

Institute of Computing, University of Campinas
Campinas, Brazil
zanoni@ic.unicamp.br

RESUMO

Dado um grafo Euleriano G , no problema de Decomposição Máxima de Ciclos em Grafos Eulerianos, estamos interessados em encontrar uma coleção de ciclos aresta-disjuntos $\{E_1, E_2, \dots, E_k\}$ em G tal que todas as arestas de G pertençam a exatamente um ciclo e k seja máximo. Apresentamos um algoritmo para resolver o problema de *pricing* da geração de colunas para um modelo de Programação Linear Inteira (PLI) apresentado por [Lancia e Serafini, 2016]. Além disso, propomos uma heurística gulosa, que busca ciclos de tamanho mínimo partindo de um vértice aleatório, e uma heurística baseada na resolução parcial do modelo PLI. Realizamos testes comparando as três abordagens em relação à qualidade das soluções e ao tempo de execução, usando conjuntos distintos de grafos Eulerianos aleatórios agrupados de acordo com a quantidade de vértices e arestas. Nossos resultados experimentais mostram que a heurística baseada no modelo PLI supera os outros métodos.

PALAVRAS CHAVE. Decomposição de Ciclos. Geração de Colunas. Programação Linear Inteira.

Tópicos: Teoria e Algoritmos em Grafos, Metaheurísticas, Otimização Combinatória.

ABSTRACT

Given an Eulerian graph G , in the Maximum Eulerian Cycle Decomposition problem, we are interested in finding a collection of edge-disjoint cycles $\{E_1, E_2, \dots, E_k\}$ in G such that all edges of G are in exactly one cycle and k is maximum. We present an algorithm to solve the pricing problem of a column generation Integer Linear Programming (ILP) model introduced by [Lancia and Serafini, 2016]. Furthermore, we propose a greedy heuristic, which searches for minimum size cycles starting from a random vertex, and a heuristic based on partially solving the ILP model. We performed tests comparing the three approaches in relation to the quality of solutions and execution time, using distinct sets of Eulerian graphs, each set grouping graphs with different numbers of vertices and edges. Our experimental results show that the ILP based heuristic outperforms the other methods.

KEYWORDS. Cycle Decomposition. Column Generation. Integer Linear Programming.

Paper Topics: Theory and Algorithms in Graphs, Metaheuristics, Combinatorial Optimization.

1. Introduction

The problem of Maximum Graph Decomposition consists of partitioning the set of edges of a graph into a collection of subsets of edges $\{E_1, E_2, \dots, E_k\}$, such that k is maximum and each subset E_i induces a specific type of subgraph (e.g., trails, cycles). This is a well studied problem in combinatorial optimization [Ganesamurthy and Paulraja, 2018; Caprara, 1999; Mynhardt and van Bommel, 2016; Heinrich, 1992; Rodger, 1990].

In this work, we deal with the version of the problem where each subset induces a cycle. To guarantee that a decomposition always exists, we assume that the input graph is always Eulerian. In this way, we call this problem the *Maximum Eulerian Cycle Decomposition* (MAX-ECD). [Caprara, 1999] showed that MAX-ECD is NP-hard. [Caprara et al., 2003] presented an ILP model with an exponential number of variables. To cope with this, [Lancia and Serafini, 2016] proposed a column generation approach to solve the ILP model of [Caprara et al., 2003]. Below we discuss our motivation to study the MAX-ECD.

In comparative genomics, the genome rearrangement distance problem receives two genomes \mathcal{G}_1 and \mathcal{G}_2 as input and consists in finding a minimum length sequence of rearrangements (mutations that alter segments of a genome) that transforms \mathcal{G}_1 into \mathcal{G}_2 . One common structure used to represent genomes is the Breakpoint Graph [Hannenhalli and Pevzner, 1999; Bafna and Pevzner, 1995]. In the Breakpoint Graph, there exists one vertex for each gene of \mathcal{G}_1 . For each pair of adjacent elements in \mathcal{G}_1 that are not adjacent in \mathcal{G}_2 , there exists a black edge connecting these elements, and for each pair of adjacent elements in \mathcal{G}_2 that are not adjacent in \mathcal{G}_1 , there exists a gray edge connecting these elements.

Most algorithms for the genome rearrangement distance need the decomposition of a Breakpoint Graph into cycles of alternating colors [Bafna and Pevzner, 1995; Hannenhalli and Pevzner, 1999; Rahman et al., 2008]. [Pinheiro et al., 2020] developed heuristics for finding decompositions of Breakpoint Graphs based on the Tabu Search metaheuristic [Glover, 1989, 1990], and presented experimental results on simulated data. [Caprara, 1999] showed a polynomial-time reduction from the Maximum Breakpoint Graph Decomposition problem to the Maximum Eulerian Cycle Decomposition problem. Therefore, the results presented here have direct application in the genome rearrangement distance problem.

In this work, we present two heuristics for the MAX-ECD. The first one is a greedy algorithm that, at each step, finds a minimum sized cycle starting at a random vertex and adds it to the current

decomposition. The second heuristic is based on solving a version of the ILP model of [Caprara et al., 2003], whose original size is exponential in the number of edges, by restricting it to only a few variables. We evaluate the performance of these algorithms in practice comparing them with the column generation presented by [Lancia and Serafini, 2016].

This paper is organized as follows. Section 2 formally introduces the problem and the definitions used in the text. Section 3 presents an integer linear programming (ILP) model for the problem. Section 4 describes the two heuristics we developed and Section 5 discusses the experimental results of the ILP model and the heuristics. Finally, in Section 6 we highlight our conclusions and give directions for future investigations.

2. Maximum Eulerian Cycle Decomposition

A graph G is an ordered pair (V, E) , where V is a set of vertices and E is a set of edges (unordered pairs of vertices). Given two vertices $u, v \in V$, we say that they are *adjacent* (or *neighbors*) if $(u, v) \in E$. In this work, we consider only simple graphs, i.e., graphs such that there are no duplicated edges and with $u \neq v$, for all edges $(u, v) \in E$.

We say that an edge (u, v) is *incident* to the vertices u and v . The *degree* of a vertex v is equal to the number of edges incident to v and it is denoted by $d(v)$.

A *trail* in a graph $G = (V, E)$ is a sequence of vertices (v_1, \dots, v_n) , such that $(v_i, v_{i+1}) \in E$, for all $1 \leq i < n$, and there are no repetitions of edges (consecutive pairs of vertices in the trail). The *length* of a trail is equal to the number of edges in it. A *circuit* in a graph $G = (V, E)$ is a trail (v_1, v_2, \dots, v_n) such that $v_1 = v_n$. A *cycle* is a circuit without repeated vertices. A graph is called *connected* if, for all pair of vertices u and v , there exists a trail from u to v .

An *Eulerian circuit* in a graph $G = (V, E)$ is a circuit that contains all edges of G . An *Eulerian graph* is a graph that has an Eulerian circuit.

A *decomposition* of a graph $G = (V, E)$ is a collection of subsets of edges $\{E_1, E_2, \dots, E_k\}$, such that the subsets are disjoint and their union is equal to E . If every subset of the collection induces a cycle in G , the partition forms a *cycle decomposition* of G , as illustrated in Figure 1.

Now, given an Eulerian graph $G = (V, E)$, in the Maximum Eulerian Cycle Decomposition problem MAX-ECD, we seek a cycle decomposition of G with maximum size. Note that, since the problem is defined only for Eulerian graphs, the existence of a cycle decomposition is guaranteed.

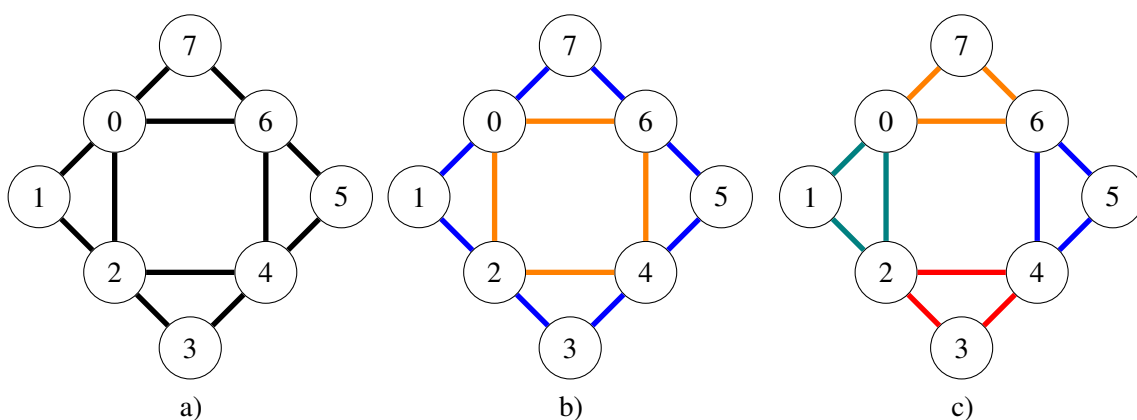


Figure 1: (a) Example of an Eulerian graph G with 8 vertices. (b) A cycle decomposition of G into two cycles $C_1 = (0, 2, 4, 6)$ and $C_2 = (0, 1, 2, 3, 4, 5, 6, 7)$. (c) A cycle decomposition of G into four cycles $C'_1 = (0, 1, 2)$, $C_2 = (2, 3, 4)$, $C_3 = (4, 5, 6)$, and $C_4 = (0, 6, 7)$, which is a decomposition of maximum cardinality.

Given a breadth-first tree $T = (V, E')$ of a graph $G = (V, E)$ rooted at $v \in V$, a *fundamental cycle* of T is any cycle that is created by inserting an edge $e \in E \setminus E'$ in T . This concept is used in the heuristics presented in Section 4.

3. Integer Linear Programming Model

Before showing the ILP model presented by [Caprara et al., 2003], we first introduce some additional notations. We use \mathcal{H} to represent the collection of cycles related to the solution constructed by the ILP model. Given an Eulerian graph $G = (V, E)$, let \mathcal{C} be the set of all cycles in G . For each cycle $C \in \mathcal{C}$, let x_C be a binary variable such that $x_C = 1$, if and only if $C \in \mathcal{H}$. Then, the model presented by [Caprara et al., 2003] reads:

$$\max \sum_{C \in \mathcal{C}} x_C \quad (1)$$

subject to:

$$\sum_{C \in \mathcal{C}: e \in C} x_C \leq 1, \quad \forall e \in E \quad (2)$$

$$x_C \in \{0, 1\}, \forall C \in \mathcal{C} \quad (3)$$

Constraint (2) guarantees that the cycles of the solution are edge-disjoint, and equation (3) defines the variables x_C as binary. Note that a set of edge-disjoint cycles that do not cover all edges in the graph is a feasible solution of this model. However, such solution can not be optimal. Indeed, as the objective function (1) maximizes the number of cycles and G is Eulerian, in any optimal solution every edge must be covered by a cycle in \mathcal{H} .

Since the size of \mathcal{C} grows exponentially with the size of E , it is not viable to create all variables of the model for large instances. Therefore, we resort to column generation and branch-and-price techniques. At the beginning of the branch-and-bound algorithm the model is fed a small subset of x_C variables and more variables are added as needed until an optimal solution of the linear relaxation is reached. The decision whether a variable is brought to the model or not is made by computing the *pricing subproblem*. For this, consider the dual of the linear relaxation of the formulation given by equations (1)–(3):

$$\min \sum_{e \in E} y_e \quad (4)$$

subject to:

$$\sum_{e \in C} y_e \geq 1, \forall C \in \mathcal{C} \quad (5)$$

$$y_e \geq 0, \forall e \in E \quad (6)$$

The pricing subproblem amounts to decide if there exists a cycle C whose associated constraint in (5) is violated by the current dual variables y^* . Since the size of \mathcal{C} is exponential in $|E|$, it is impractical to check all such constraints. Instead, we use a combinatorial algorithm described below to solve the pricing subproblem.

For each cycle $C \in \mathcal{C}$, constraints (5) require that the sum of variables y_e^* , for $e \in C$, to be greater than or equal to 1. Therefore, given a cycle C , if $\sum_{e \in C} y_e^* < 1$, the associated constraint is violated. Hence, in the separation algorithm, we create a weighted graph G' with the same vertices

Algorithm 1: Pseudocode for finding the minimum weight cycle.

Input: A graph $G' = (V, E)$ and $weight$
Output: The cycle found ($cycle$) and its weight ($minimum$)

- 1 $minimum \leftarrow \infty$
- 2 $cycle \leftarrow \emptyset$
- 3 **foreach** $e = (u, v) \in E$ **do**
- 4 $weight' \leftarrow weight$
- 5 $weight'[e] \leftarrow \infty$
- 6 $(w, P) \leftarrow minimum_weight_path(u, v, G', weight')$
- 7 $w \leftarrow w + weight[e]$
- 8 $C \leftarrow P \cup \{e\}$
- 9 **if** $w < minimum$ **then**
- 10 $minimum \leftarrow w$
- 11 $cycle \leftarrow C$
- 12 **return** ($minimum, cycle$)

and edges of G , where each edge e has weight y_e^* . A constraint of the dual model is violated if, and only if, the minimum weight cycle of G' has weight less than one. Thus, the algorithm finds a minimum weight cycle C^* in G' and, if its weight is less than one, the constraint related to C^* is violated and the variable x_{C^*} is added to the primal model. If the weight of C^* is greater than or equal to one, no variables need to be added to the linear relaxation of the primal model [Lancia and Serafini, 2016], since optimality has been reached.

So, the pricing problem is to find a minimum weight cycle C^* in the weighted graph G' , without considering cycles related to variables x_C that were fixed to zero by the branch-and-bound algorithm, since we want to find new cycles to consider in the primal model. When no variables x_C are fixed to zero, we can use the MINIMUM-CYCLE algorithm described in Algorithm 1. For each edge (u, v) , the algorithm finds a minimum weight path P from u to v , which does not have the edge (u, v) , and creates the cycle $P \cup (u, v)$. Note that this is the minimum weight cycle that contains the edge (u, v) . Then, the algorithm returns the generated cycle with minimum weight.

When the variables $\{x_{C_1}, x_{C_2}, \dots, x_{C_k}\}$ are fixed to zero, we need to forbid these cycles in the pricing problem. For this, we create copies G_{z_1, z_2, \dots, z_k} of G' , such that $1 \leq z_i \leq |C_i|$ for $1 \leq i \leq k$. In each copy G_{z_1, z_2, \dots, z_k} , we forbid the cycles C_1 to C_k by updating the weight of the z_i -th edge of C_i to ∞ . The algorithm MINIMUM-CYCLE is executed for each altered copy of G' and the minimum weight cycle found in all these executions is the solution of the pricing problem.

4. Heuristics for MAX-ECD

In this section, we present two heuristics developed to solve the problem, which are named GREEDY and ILP-HEURISTIC.

The GREEDY heuristic finds a cycle decomposition as described in Algorithm 2. At each iteration, a random vertex v with degree $d(v) > 0$ is chosen. The algorithm then performs a breadth-first search to find a minimum size cycle C starting at v , adds this cycle to the decomposition, and removes all edges of C from the graph.

Since the edges of a cycle added to the solution \mathcal{H} are removed before starting a new iteration, all cycles in \mathcal{H} are edge-disjoint. Note that after removing all edges of a cycle from the graph, all vertices remain with an even degree. Therefore, at each iteration, any vertex v with $d(v) > 0$ is in at least one cycle. Furthermore, since the algorithm always finds a cycle, eventually

$|E| = 0$, and the solution covers all edges of the original graph. The GREEDY heuristic has a time complexity of $O(|E|^2)$.

Algorithm 2: Pseudocode for the **GREEDY** heuristic.

Input: A graph $G = (V, E)$
Output: A cycle decomposition \mathcal{H}

- 1 $\mathcal{H} = \emptyset$
- 2 **while** $|E| > 0$ **do**
- 3 choose a vertex v at random, such that $d(v) > 0$
- 4 perform a breadth-first search to find a fundamental cycle C of v of minimum size
- 5 add C to \mathcal{H}
- 6 remove the edges of C from G
- 7 **return** \mathcal{H}

The ILP-HEURISTIC consists in the following. Given a set of cycles $\mathcal{C}' \subset \mathcal{C}$, we create the model from Section 3 using only variables x_C such that $C \in \mathcal{C}'$. To ensure that the modified model is feasible, we can add the cycles of at least one valid decomposition in \mathcal{C}' . Given a parameter k , we create the set \mathcal{C}' with the cycles of k random decompositions created by the GREEDY heuristic.

This strategy guarantees that the optimal solution of the ILP-HEURISTIC is better than or equal to the best of the k random solutions generated by GREEDY heuristic. However, a solution from the ILP-HEURISTIC may not cover all edges of the input graph, since we run the ILP model using a subset of \mathcal{C} . Therefore, given a solution \mathcal{H} from the ILP-HEURISTIC for an Eulerian graph $G = (V, E)$, we apply a post-processing step with the GREEDY heuristic that finds a decomposition \mathcal{H}' for the graph $G' = (V, E')$, where $E' = \{(u, v) \in E \mid (u, v) \notin \mathcal{H}\}$, and returns $\mathcal{H} \cup \mathcal{H}'$, which is then a valid decomposition for G . When we refer to the ILP-HEURISTIC, we assume that the post-processing step is also performed.

5. Experimental Analysis

In this section, we describe the experimental results as well as the instances generation and improvements on the execution of the ILP model.

The experiments for the MAX-ECD were done in different sets of random Eulerian graphs, each set grouping graphs with different numbers of vertices and edges. To create these sets, we used the algorithm defined by [Hakimi, 1962], which is described next. Given a sequence of n positive integers, the algorithm creates a graph G with n vertices such that there exists a one-to-one relationship between the numbers of the sequence and the degree of each vertex v in G , or the algorithm determines that such graph does not exist. Note that the list of numbers given as a parameter to Hakimi's algorithm must have only even numbers, since we want Eulerian graphs. To generate this sequence we use Algorithm 3, that receives two parameters m and n , such that $m \geq n$, and it returns a list d_1, d_2, \dots, d_n such that $\sum_{i=1}^n d_i = 2m$ and $d_i = 2k$, for any $1 \leq i \leq n$ and $k \in \mathbb{Z}^+$.

Recall that Hakimi's algorithm may construct a disconnected graph. In this case, the following procedure is applied while the number of connected components of G is greater than 1. Choose two connected components C and C' ; then, choose at random two edges $(u, v) \in C$ and $(x, y) \in C'$; at last, replace them by the edges (u, x) and (v, y) , which decreases the number of connected components of the graph by one. Note that this procedure does not modify the degree of the vertices, and ensures that the resulting graph is connected.

For each combination of parameters $n \in \{10, 20, \dots, 90, 100\}$ and $m \in \{0.1, 0.2, 0.3, 0.4,$

Algorithm 3: Pseudocode for the creation of a random sequence used by the algorithm of Hakimi.

Input: Parameters n and m
Output: A list of n integers S

- 1 let S be a list of n integers of value 2
- 2 let $m' \leftarrow m - n$
- 3 **while** $m' > 0$ **do**
- 4 let i be a value from 1 to n chosen uniformly at random
- 5 $S_i \leftarrow S_i + 2$
- 6 $m' \leftarrow m' - 1$
- 7 **return** S

0.5}, we created a set containing 20 Eulerian graphs with n vertices and $\lfloor m \binom{n}{2} \rfloor$ edges¹.

5.1. Experimental Results

The best trade-off between time and quality of solution for the GREEDY heuristic was obtained when $k = 100$ (i.e., the algorithm was executed 100 times for each instance). Tables 1a-1e show the average number of cycles returned when considering the five different densities, and Table 1f groups these results with all densities.

In Table 1, each entry at column `G_avg` corresponds to the average of 100 solutions provided by the GREEDY for each instance in the corresponding set. Column `G_max` indicates the average considering only the best decomposition returned by GREEDY for each instance.

We tested the ILP model with column generation described in Section 3 with different sets of initial variables and primal bounds. We tested the following sets as initial variables: all triangles (i.e., cycles with three vertices) from the graph; all fundamental cycles, which includes all triangles; and the set of cycles returned from the ILP-HEURISTIC.

We also tested the ILP model with no initial cycles, and this configuration returned the best solutions on average, whose results are in column `ILP_cg` of Table 1. Column `OPT` on its right indicates the proportion of instances where the algorithm has found a proven optimal solution.

We also tested to warm start the branch-and-price algorithm with the primal bounds obtained by the heuristics GREEDY and ILP-HEURISTIC. Although, in this case, the solver was able to find a feasible solution for all instances, we did not observe a relevant gain in performance. Even for instances where `ILP_cg` had already found a solution, few or no improvements at all were observed with the strategies mentioned above. Thus, the results reported in the column `ILP_cg` of the Table 1 correspond to the execution using no warm start.

Column `ILP_h` shows the average results returned by ILP-HEURISTIC described in Section 3, using the 100 solutions returned by GREEDY, and column `OPT` on its right indicates the percentage of instances where the algorithm found an optimal solution, proved by `ILP_cg_h`, an ILP model with column generation using the solution found by ILP-HEURISTIC as the primal bound and the cycles returned by GREEDY as initial variables. For column `ILP_cg`, there is a “-” when the ILP model was not able to finish execution and find a feasible solution for any of the instances of the set.

In most cases, the ILP-HEURISTIC was able to considerably improve the 100 random decompositions found by the GREEDY heuristic. The ILP-HEURISTIC also gave better results

¹Instances are available at: <https://github.com/compbiogroup/Maximum-Eulerian-Cycle-Decomposition-Problem>

when compared with the ILP model with column generation. Note that we use the bounds of the ILP model to prove that a solution is optimal and, for large values of n and m , these bounds get worse, which could also be a reason for the low number of optimal solutions found by the ILP-HEURISTIC as the value of n and m increases.

Table 1: Average number of cycles of the solutions of the following approaches: average of 100 executions of GREEDY (G_avg); maximum of 100 executions of GREEDY (G_max); ILP model with column generation (ILP_cg); and ILP-HEURISTIC (ILP_h). Column OPT next to ILP_cg has the percentage of instances that the ILP model with column generation found an optimal solution; column OPT next to ILP_h has the percentage of instances that the ILP-HEURISTIC found an optimal solution (proved by ILP_cg_h).

n	G_avg	G_max	ILP_cg	OPT	ILP_h	OPT	n	G_avg	G_max	ILP_cg	OPT	ILP_h	OPT
10	1.00	1.00	1.00	100%	1.00	100%	10	1.00	1.00	1.00	100%	1.00	100%
20	1.00	1.00	1.00	100%	1.00	100%	20	8.36	9.10	9.10	100%	9.10	100%
30	6.83	7.60	7.60	100%	7.60	100%	30	20.90	22.90	23.85	95%	23.85	100%
40	14.46	16.55	16.85	95%	16.85	100%	40	39.01	41.80	44.65	65%	44.75	90%
50	24.39	27.20	28.80	80%	28.85	100%	50	63.25	66.60	71.95	20%	72.80	80%
60	37.48	40.95	43.90	40%	44.15	90%	60	93.49	97.05	105.35	5%	107.20	60%
70	52.51	56.15	61.25	15%	61.85	40%	70	130.26	134.45	-	-	148.40	0%
80	71.19	75.25	81.95	0%	83.35	10%	80	172.83	177.45	-	-	195.50	0%
90	92.22	97.00	-	-	107.80	0%	90	221.70	227.05	-	-	248.50	0%
100	116.43	121.20	-	-	134.75	0%	100	276.52	282.20	-	-	304.25	0%

a) Results for $m = 10\%$.

b) Results for $m = 20\%$.

n	G_avg	G_max	ILP_cg	OPT	ILP_h	OPT	n	G_avg	G_max	ILP_cg	OPT	ILP_h	OPT
10	2.89	3.00	3.00	100%	3.00	100%	10	4.83	5.05	5.05	100%	5.05	100%
20	14.63	16.00	16.35	100%	16.35	100%	20	21.04	22.70	23.65	85%	23.70	100%
30	34.87	37.15	39.55	80%	39.65	100%	30	49.51	52.00	55.95	80%	56.10	100%
40	64.74	67.75	73.65	50%	73.80	90%	40	90.54	93.75	101.30	50%	102.05	100%
50	103.20	107.10	115.55	20%	117.00	75%	50	144.30	148.65	-	-	162.35	100%
60	151.98	156.35	-	-	172.65	60%	60	210.99	216.65	-	-	235.45	100%
70	209.63	215.05	-	-	237.15	30%	70	290.41	297.95	-	-	321.75	100%
80	277.00	283.35	-	-	308.80	0%	80	382.82	393.40	-	-	420.95	95%
90	354.29	363.20	-	-	394.65	0%	90	487.99	500.85	-	-	533.25	55%
100	441.10	450.90	-	-	487.75	5%	100	606.16	622.10	-	-	658.35	10%

c) Results for $m = 30\%$.

d) Results for $m = 40\%$.

n	G_avg	G_max	ILP_cg	OPT	ILP_h	OPT	n	G_avg	G_max	ILP_cg	OPT	ILP_h	OPT
10	6.05	6.40	6.40	100%	6.40	100%	10	3.15	3.29	3.29	100%	3.29	100%
20	27.39	29.20	30.75	100%	30.75	100%	20	14.48	15.60	16.17	97%	16.18	100%
30	63.88	66.50	71.20	60%	71.75	100%	30	35.20	37.23	39.63	83%	39.79	100%
40	116.56	120.30	128.55	35%	129.70	100%	40	65.06	68.03	73.00	59%	73.43	95%
50	185.03	190.50	-	-	203.95	100%	50	104.03	108.01	-	-	116.99	91%
60	269.97	277.30	-	-	295.00	100%	60	152.78	157.66	-	-	170.89	82%
70	370.75	380.45	-	-	402.00	100%	70	210.71	216.81	-	-	234.23	53%
80	488.34	500.95	-	-	526.00	100%	80	278.44	286.08	-	-	306.92	41%
90	621.70	637.95	-	-	666.75	75%	90	355.58	365.21	-	-	390.19	26%
100	771.67	791.85	-	-	823.25	30%	100	442.37	453.65	-	-	481.67	9%

e) Results for $m = 50\%$.

f) Results for all values of m .

Figure 2 shows the percentage of optimal solutions found by the ILP model with column generation (ILP_cg) and by the ILP-HEURISTIC. We can see that the ILP-HEURISTIC was significantly better on finding optimal solutions in this experiment, finding optimal solutions in

	10	20	30	40	50	60	70	80	90	100	V
10%	100%	100%	100%	95%	80%	40%	15%	0%	0%	0%	53%
20%	100%	100%	95%	65%	20%	5%	0%	0%	0%	0%	38%
30%	100%	100%	80%	50%	20%	0%	0%	0%	0%	0%	35%
40%	100%	85%	80%	50%	10%	0%	0%	0%	0%	0%	32%
50%	100%	100%	60%	35%	0%	0%	0%	0%	0%	0%	30%
D	100%	97%	83%	59%	26%	9%	3%	0%	0%	0%	38%

a) ILP model with column generation (ILP_cg)

	10	20	30	40	50	60	70	80	90	100	V
10%	100%	100%	100%	100%	100%	90%	35%	10%	0%	0%	64%
20%	100%	100%	100%	90%	80%	60%	0%	0%	0%	0%	53%
30%	100%	100%	100%	85%	75%	60%	30%	0%	0%	5%	56%
40%	100%	100%	100%	100%	100%	100%	100%	95%	55%	10%	86%
50%	100%	100%	100%	100%	100%	100%	100%	100%	75%	30%	90%
D	100%	100%	100%	95%	91%	82%	53%	41%	26%	9%	70%

b) ILP-HEURISTIC (ILP_h)

Figure 2: The horizontal axis indicates the number of vertices of the instances, and the vertical axis indicates the density of the graph. Each cell indicates (a) the percentage of optimal solutions found by the ILP model with column generation and (b) the percentage of optimal solutions found by the ILP-HEURISTIC for each set of instances. Line D and column V show the percentage of optimal solutions grouped by density and by number of vertices, respectively.

69.7% of the instances, while ILP_cg found the optimal solution only on 37.7% of them. We can observe that the number of optimal solutions from ILP-HEURISTIC increases for graphs with a higher number of edges. This can be explained by the fact that the denser the graphs, the greater the chances of having multiple cycles of size 3 (called triangles). Table 2 shows the percentage of triangles in the solutions returned by the ILP-HEURISTIC, where we can observe a clear trend of this type of cycle as the density increases. For instance, while the percentage of triangles is 29.96% when $m = 10\%$, this percentage grows to 95.40% when $m = 50\%$.

Tables 3a-3e show the average time (in seconds) spent by the algorithms when considering the five different densities, and Table 3f groups these results with all densities. The time limit was set to 1800 seconds for every algorithm.

In the cases ILP_cg_h was able to prove the optimality, this was done in 20 seconds on average, and in only 1.9% of cases it needed more than 60 seconds. In the remaining cases, ILP_cg_h was not able to prove optimality nor to improve the solution obtained by ILP-HEURISTIC, within the limit of 1800 seconds.

Among the cases where it was not possible to prove optimality for the solution obtained by the ILP-HEURISTIC, in 68.0% of them neither ILP-HEURISTIC nor ILP_cg_h managed to

Table 2: Percentage of the number of cycles of size 3 (triangles) in the solutions returned by the ILP-HEURISTIC.

n	$m = 10\%$	$m = 20\%$	$m = 30\%$	$m = 40\%$	$m = 50\%$
10	0.00%	0.00%	31.67%	59.41%	67.19%
20	0.00%	36.81%	63.30%	81.43%	91.87%
30	24.34%	53.46%	75.13%	89.19%	97.63%
40	27.89%	60.65%	83.40%	94.27%	99.31%
50	34.09%	67.29%	86.40%	98.18%	99.93%
60	37.41%	72.42%	92.56%	99.41%	100.00%
70	38.54%	75.57%	94.89%	99.77%	99.75%
80	42.74%	77.85%	95.34%	99.30%	99.33%
90	46.10%	81.34%	97.37%	99.45%	99.45%
100	48.53%	83.44%	98.45%	99.45%	99.50%

obtain an optimal solution within the limit of 1800 seconds. In the other 32.0%, ILP-HEURISTIC found an optimal solution (considering the decompositions given as input), but ILP_cg_h was unable to prove optimality within the time limit.

All the experiments were conducted on a PC equipped with a 2.4GHz Intel Core™ i7-8700T, with 12 cores and 8 GB of RAM, running Ubuntu 18.04.3. We used SCIP 6.0.2 [Achterberg, 2009] as the ILP solver and Gurobi 8.1.1 [Gurobi Optimization, LLC, 2020] as the LP solver. Since the ILP-HEURISTIC does not need column generation, we used Gurobi as the ILP solver for this heuristic.

6. Conclusion

In this work we presented algorithms for the Maximum Eulerian Cycle Decomposition problem, alongside an experimental study of these algorithms and a column generation ILP model for the problem.

Our experiments showed good results for the ILP heuristic, which is an heuristic that partially solves the ILP model activating only the variables related to good cycles found by a greedy heuristic.

For future work, we intend to propose new algorithms based on metaheuristics used in the literature for combinatorial problems, such as Genetic Algorithms (GA) and Greedy Randomized Adaptive Search Procedure (GRASP).

Acknowledgments. This work was supported by the National Council of Technological and Scientific Development, CNPq (425340/2016-3 and 304380/2018-0), the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Finance Code 001, and the São Paulo Research Foundation, FAPESP (grants 2013/08293-7, 2015/11937-9, 2017/12646-3, 2019/25410-3, and 2019/27331-3).

Table 3: Average running time in seconds of the following approaches: 100 executions of GREEDY (G_avg); ILP-HEURISTIC (ILP_h); ILP model with column generation (ILP_cg); and the ILP model with column generation using the solution found by ILP-HEURISTIC and the cycles returned by GREEDY (ILP_cg_h).

n	G_avg	ILP_h	ILP_cg	ILP_cg_h	n	G_avg	ILP_h	ILP_cg	ILP_cg_h
10	0.00	0.00	0.00	0.00	10	0.00	0.00	0.00	0.00
20	0.01	0.00	0.00	0.00	20	0.08	0.00	0.00	0.00
30	0.21	0.00	0.00	0.00	30	0.55	0.06	116.90	0.03
40	1.02	0.02	90.05	0.06	40	1.94	0.63	635.40	230.08
50	3.89	0.23	391.90	0.20	50	5.31	4.57	1525.60	369.83
60	6.97	1.55	1082.40	218.03	60	12.09	113.80	1786.70	722.72
70	11.42	6.20	1629.85	1189.97	70	23.06	826.49	1800.00	1800.00
80	17.28	82.59	1800.00	1620.33	80	41.02	1709.79	1800.00	1800.00
90	35.56	752.49	1800.00	1800.00	90	68.08	1800.00	1800.00	1800.00
100	48.12	1800.00	1800.00	1800.00	100	107.77	1800.00	1800.00	1800.00

a) Results for $m = 10\%$.

b) Results for $m = 20\%$.

n	G_avg	ILP_h	ILP_cg	ILP_cg_h	n	G_avg	ILP_h	ILP_cg	ILP_cg_h
10	0.01	0.00	0.00	0.00	10	0.01	0.00	0.00	0.00
20	0.15	0.01	0.00	0.01	20	0.21	0.02	270.00	3.84
30	0.83	0.22	373.80	62.96	30	1.31	0.69	415.10	28.32
40	3.13	2.24	1071.05	360.14	40	5.31	5.41	1054.95	0.29
50	9.21	24.01	1558.15	494.85	50	15.63	19.72	1697.00	0.98
60	20.90	594.01	1800.00	906.69	60	37.36	146.00	1800.00	20.71
70	42.67	1432.95	1800.00	1261.47	70	77.82	107.12	1800.00	4.51
80	79.32	1800.00	1800.00	1800.00	80	150.90	620.56	1800.00	99.14
90	139.11	1800.00	1800.00	1800.00	90	268.44	998.60	1800.00	822.75
100	230.98	1749.13	1800.00	1710.90	100	454.00	1686.91	1800.00	1623.69

c) Results for $m = 30\%$.

d) Results for $m = 40\%$.

n	G_avg	ILP_h	ILP_cg	ILP_cg_h	n	G_avg	ILP_h	ILP_cg	ILP_cg_h
10	0.01	0.00	0.00	0.00	10	0.01	0.00	0.00	0.00
20	0.30	0.03	4.75	0.02	20	0.15	0.02	54.95	0.77
30	2.05	0.29	742.80	0.10	30	0.99	0.25	329.72	18.28
40	8.61	3.37	1243.40	0.37	40	4.00	2.33	818.97	118.19
50	25.35	17.44	1800.00	1.14	50	11.88	13.20	1394.53	173.40
60	62.28	69.24	1800.00	2.65	60	27.92	184.92	1653.82	374.16
70	134.90	325.33	1800.00	6.01	70	57.97	539.62	1765.97	852.39
80	265.77	1266.17	1800.00	14.29	80	110.86	1095.82	1800.00	1066.75
90	487.29	1428.34	1800.00	474.75	90	199.70	1355.89	1800.00	1339.50
100	902.33	1665.81	1800.00	1273.16	100	348.64	1740.37	1800.00	1641.55

e) Results for $m = 50\%$.

f) Results for all values of m .

References

- Achterberg, T. (2009). SCIP: Solving Constraint Integer Programs. *Mathematical Programming Computation*, 1(1):1–41.
- Bafna, V. and Pevzner, P. A. (1995). Sorting Permutations by Transpositions. In *Proceedings of the 6th Annual ACM-SIAM Symposium on Discrete Algorithms (SODA'1995)*, p. 614–623, Philadelphia, PA, USA. Society for Industrial and Applied Mathematics.
- Caprara, A. (1999). Sorting Permutations by Reversals and Eulerian Cycle Decompositions. *SIAM Journal on Discrete Mathematics*, 12(1):91–110.
- Caprara, A., Panconesi, A., and Rizzi, R. (2003). Packing Cycles in Undirected Graphs. *Journal of Algorithms*, 48(1):239–256.
- Ganesamurthy, S. and Paulraja, P. (2018). 2p-Cycle Decompositions of Some Regular Graphs and Digraphs. *Discrete Mathematics*, 341(8):2197–2210.
- Glover, F. W. (1989). Tabu search - part I. *INFORMS Journal on Computing*, 1(3):190–206.
- Glover, F. W. (1990). Tabu search - part II. *INFORMS Journal on Computing*, 2(1):4–32.
- Gurobi Optimization, LLC (2020). Gurobi Optimizer Reference Manual. <https://www.gurobi.com>. Access in November 23, 2020.
- Hakimi, S. L. (1962). On Realizability of a Set of Integers as Degrees of the Vertices of a Linear Graph. I. *Journal of the Society for Industrial and Applied Mathematics*, 10(3):496–506.
- Hannenhalli, S. and Pevzner, P. A. (1999). Transforming Cabbage into Turnip: Polynomial Algorithm for Sorting Signed Permutations by Reversals. *Journal of the ACM*, 46(1):1–27.
- Heinrich, K. (1992). Path Decomposition. *Le Matematiche*, 47(2):241–258.
- Lancia, G. and Serafini, P. (2016). Deriving Compact Extended Formulations via LP-based Separation Techniques. *Annals of Operations Research*, 240(1):321–350.
- Mynhardt, C. M. and van Bommel, C. M. (2016). Triangle Decompositions of Planar Graphs. *Discussiones Mathematicae Graph Theory*, 36(3):643–659.
- Pinheiro, P. O., Alexandrino, A. O., Oliveira, A. R., de Souza, C. C., and Dias, Z. (2020). Heuristics for Breakpoint Graph Decomposition with Applications in Genome Rearrangement Problems. In *Proceedings of the 13th Brazilian Symposium on Bioinformatics (BSB'2020)*, p. 129–140. Springer International Publishing.
- Rahman, A., Shatabda, S., and Hasan, M. (2008). An Approximation Algorithm for Sorting by Reversals and Transpositions. *Journal of Discrete Algorithms*, 6(3):449–457.
- Rodger, C. A. (1990). Graph Decomposition. *Le Matematiche*, 45(1):119–140.