THE WEIGHTED SITTING CLOSER TO FRIENDS THAN ENEMIES PROBLEM IN THE LINE*

JULIO ARACENA † AND CHRISTOPHER THRAVES CARO‡

Abstract. The weighted *Sitting Closer to Friends than Enemies* (SCFE) problem is to find an injection of the vertex set of a given weighted graph into a given metric space so that, for every pair of incident edges with different weight, the end vertices of the heavier edge are closer than the end vertices of the lighter edge. In this work, we provide a characterization of the set of weighted graphs with an injection in $\mathbb{R}$ that satisfies the restrictions of the weighted SCFE problem. Indeed, given a weighted graph G, we define a polyhedron $M(G)\mathbf{x} \leq \mathbf{b}$, and show that a weighted graph G has an injection that solves the weighted SCFE problem in $\mathbb{R}$ if and only if $M(G)\mathbf{x} \leq \mathbf{b}$ is not empty. As a consequence of this result, we conclude that deciding the existence of, and constructing such an injection for a given *complete* weighted graph can be done in polynomial time. On the other hand, we show that deciding if an *incomplete* weighted graph has such an injection in $\mathbb{R}$ is NP-Complete. Nevertheless, we prove that if the number of missing edges is constant, such decision can be done in polynomial time.

Key words. The SCFE Problem, Robinsonian Matrices, Valid Distance Drawings, Weighted Graphs, Metric Spaces, Seriation Problem.

AMS subject classifications. 05C22, 05C62, 05C85, 68R10

1. Introduction.

Consider a data set. The task is to construct a graphic representation of the data set so that similarities between data points are graphically expressed. To complete this task, the only information available is a *similarity matrix* of the data set, i. e., a square matrix whose entry ij contains a similarity measure between data points i and j (the larger the value the more similar the data points are). Hence, the task is to draw all data points in a *paper* so that for every three data points i, j, and k, if i is at least as similar to j than k, then i should be placed closer in the drawing to j than k. In colloquial words, for each data point j, the farther the other data points are, the less similar they are to j.

A slightly simpler version of this problem, introduced in [12], has been studied under the name of the Sitting Closer to Friends than Enemies (SCFE) problem. The SCFE problem uses signed graphs as an input. Therefore, the similarity matrix has entries 1 and -1, representing similarity and dissimilarity, or friendship and enmity between the data points, from where the problem obtains its name. The SCFE problem has been studied in the real line [12, 7, 18] and in the circumference [2] (which means that the *paper* is the real line or the circumference). In both cases, the real line and the circumference, it has been shown that deciding the existence of such an injection for a given signed graph is NP-Complete. Nevertheless, in both cases again, when the problem is restricted to complete signed graphs there exists a characterization of the families of complete signed graphs that admit a solution for the SCFE problem and it can be decided in polynomial time [12, 2]. Therefore, a natural next step is to consider now the case when similarities range in an extended set of values.

*Submitted to the editors DATE.

Funding: J. Aracena was partially supported by CONICYT-Chile through the project AFB170001 of the PIA Program: *Concurso Apoyo a Centros Científicos y Tecnológicos de Excelencia con Financiamiento Basal.*

†CI²MA and Departamento de Ingeniería Matemática, Facultad de Ciencias Físicas y Matemáticas, Universidad de Concepción, Concepción, Chile. (jaracena@ing-mat.udec.cl)

‡Departamento de Ingeniería Matemática, Facultad de Ciencias Físicas y Matemáticas, Universidad de Concepción, Concepción, Chile. (cthraves@ing-mat.udec.cl).

Here, we consider the case when similarities are restricted to be positive values, and two points are more similar if their similarity value is larger.

The SCFE problem in the line seems to be equivalent to the *Seriation* problem. Liiv in [16] defines the Seriation problem as "an exploratory data analysis technique to reorder objects into a sequence along a one-dimensional continuum so that it best reveals regularity and pattering among the whole series". Seriation has applications in archaeology [19], data visualization [3], exploratory analysis [11], bioinformatics [24], and machine learning [8], among others. Liiv in [16] presents an interesting survey on seriation, matrix reordering and its applications. The first important contribution of this document is to show that the SCFE and the Seriation problems are different. Indeed, we show that seriation is a necessary condition to solve the SCFE problem, but it is not sufficient.

To continue with our exposition, we introduce the notation and definitions used along the document in Section 2. The rest of the document is organized as follows. In Section 3, we present the state of the art and contextualize our contributions. In Section 4, we present the characterization of weighted graphs with an injection in $\mathbb{R}$ that satisfies the restrictions of the SCFE problem. Furthermore, we present the results related with complete weighted graphs. In Section 5, we present the results regarding incomplete weighted graphs. We conclude in Section 6 with some final remarks and future work.

2. Notation and Definitions. We use standard notation. A graph is denoted by $G = (V, E)$. We consider only undirected graphs, without parallel edges and loopless. The set of vertices of G is V and the set of edges is E, a set of 2-elements subsets of V. We use n and m to denote $|V|$ and $|E|$, respectively. Two distinct vertices i and j in V are said to be *neighbors* if $\{i, j\} \in E$. In that case, we say that they are connected by an edge which is denoted by $\{i, j\}$. Along the document we also use the number of missing edges. Hence, let r be the number of pairs $\{i, j\}$ such that $\{i, j\} \notin E$. It is worth noting that $m + r = n(n-1)/2$. A graph is said to be *complete* if every pair of distinct vertices is connected by an edge, otherwise, we say that it is *incomplete*.

In this document, we work with weighted graphs. We denote by $w : E \to \mathbb{R}^+$ a positive real valued function that assigns $w(\{i, j\})$, a positive real value, to the edge $\{i, j\}$ in E. We denote by L the number of different values that the function w assigns. For our purposes, we consider that w is a similarity measure, i. e., for any $\{i, j\} \in E$ the value $w(\{i, j\})$ measures how similar i and j are. Moreover, we consider that the similarity measure is symmetric, therefore, $w(\{i, j\}) = w(\{j, i\})$. We consider that the larger the similarity measure is, the more similar the vertices are. It is worth mentioning that the fact that the weights are positive is just a choice made for simplicity. Actually, the weights can also be negative and all our results will still be valid.

Let $(\mathcal{M}, d)$ be a metric space. A *drawing* of a graph $G = (V, E)$ into $\mathcal{M}$ is an injection $D : V \to \mathcal{M}$. We define a certain type of drawings that capture the requirements of the SCFE problem.

Definition 2.1. *Let $G = (V, E)$ be a graph, and $w : E \to \mathbb{R}^+$ be a positive function on E. Let $(\mathcal{M}, d)$ be a metric space. We say that a drawing D of G into $\mathcal{M}$ is* valid distance *if, for all pair $\{i, j\}$, $\{i, k\}$ of incident edges in E such that $w(\{i, j\}) > w(\{i, k\})$,*

$$d(D(i), D(j)) < d(D(i), D(k)).$$

In colloquial words, a drawing is valid distance, or simply *valid*, when it places vertices i and j strictly closer than k and j in $\mathcal{M}$ whenever i and j have a strictly larger similarity measure than k and j. Now, the weighted SCFE problem in its most general presentation is defined as follows.

DEFINITION 2.2. *Given a weighted graph G and a metric space $\mathcal{M}$, the weighted SCFE problem in $\mathcal{M}$ is to decide whether G has a valid drawing in $\mathcal{M}$, and, in case of a positive answer on the first question, find one.*

In this document, we focus our attention on the case when the metric space is the real line, i. e., we consider the metric space to be the set of real values $\mathbb{R}$ with the Euclidean distance.

Since we present a matrix oriented analysis, we introduce the next two matrix related definitions. Given a matrix A, the entry in the *i-th* row and *j-th* column of A is denoted by A_{ij}. For every weighted graph G, we denote by $A(G)$ the square matrix defined as follows:

$$A(G)_{ij} = \begin{cases} * & \text{if } i \neq j \text{ and } \{i,j\} \notin E, \\ w(\{i,j\}) & \text{if } i \neq j \text{ and } \{i,j\} \in E, \\ \max_{\{k,l\}\in E} w(\{k,l\}) & \text{if } i = j. \end{cases}$$

We call this matrix the *similarity matrix* of G also known as the extended weighted *adjacency matrix* of G. The *i-th* row (and *i-th* column) contains the similarities between vertex i and the rest of the vertices of G. We may use only A when the graph G is contextually clear. Note that the similarity matrix of any weighted graph is symmetric since $w(\{i,j\}) = w(\{j,i\})$. The similarity matrix of a complete weighted graph does not have entries with the symbol $*$. In that case, we say that the similarity matrix is *complete*, otherwise we say that it is *incomplete*.

W. S. Robinson in [22] introduced Robinsonian matrices. A complete similarity matrix is said to be in *Robinsonian form* if its entries are monotone nondecreasing in rows and columns when moving towards the diagonal, i. e., if for all $1 \leq i < j \leq n$,

$$A_{ij} \leq \min\{A_{ij-1}, A_{i+1j}\}.$$

On the other hand, a complete similarity matrix is *Robinsonian* if its rows and columns can be reordered simultaneously such that it passes to be in Robinsonian form.

Robinsonian matrix definition can be naturally extended to incomplete matrices. In that case, a similarity matrix is in *Robinsonian form* if its entries are monotone nondecreasing in rows and columns when moving towards the diagonal considering only numerical entries, i. e., if for all $1 \leq i < j < k \leq n$ such that $A_{ik} \neq *$, $A_{ij} \neq *$ and $A_{jk} \neq *$,

$$A_{ik} \leq \min\{A_{ij}, A_{jk}\}.$$

Again, we say that a similarity matrix is *Robinsonian* if its rows and columns can be simultaneously reordered in such a way that it passes to be in Robinsonian form.

3. Context, Related Work, and Our Contributions. Robinsonian matrices were defined by W. S. Robinson in [22] in a study on how to order chronologically archaeological deposits. The *Seriation* problem introduced in the same work then is to decide whether the similarity matrix of a data set is Robinsonian and write it in Robinsonian form. Recognition of complete Robinsonian matrices has been studied by several authors. Mirkin et al. in [17] presented an $O(n^4)$ recognition algorithm, where $n \times n$ is the size of the matrix. On the other hand, using divide and conquer techniques, Chepoi et al. in [4] introduced an $O(n^3)$ recognition algorithm. Later, Préa and Fortin in [20] provided an $O(n^2)$ optimal recognition algorithm for complete Robinsonian matrices using PQ trees.

Using the relationship between Robinsonian matrices and unit interval graphs presented in [21], Monique Laurent and Matteo Seminaroti in [13] introduced a recognition algorithm for Robinsonian matrices that uses Lex-BFS, whose time complexity is $O(L(r+n))$, where r is the number of zero entries in the matrix[1], and L is the number of different values in the matrix. Later in [14], the same authors presented a recognition algorithm with time complexity $O(n^2 + r \log n)$ that uses similarity first search. Again, using the relationship between Robinsonian matrices and unit interval graphs, Laurent et al. in [15] gave a characterization of Robinsonian matrices via forbidden patterns.

The Seriation problem also has been studied as an optimization problem. Given an $n \times n$ matrix D, *seriation in the presence of errors* is to find a Robinsonian matrix R that minimizes the error defined as: $\max ||D_{ij} - R_{ij}||$ over all i and j in $\{1, 2, 3, \ldots, n\}$. Chepoi et al. in [5] proved that seriation in the presence of errors is an NP-Hard problem. Later in [6], Chepoi and Seston gave a factor 16 approximation algorithm. Fortin in [9] surveyed the challenges for Robinsonian matrix recognition.

The SCFE problem was first introduced by Kermarrec and Thraves in [12]. Besides the introduction of the SCFE problem, the authors of [12] also characterized the set of complete signed graphs with a valid drawing in $\mathbb{R}$ and presented a polynomial time recognition algorithm. Later, Cygan et al. in [7] proved that the SCFE problem is NP-Complete if it is not restricted to complete signed graphs. Moreover, they gave a different characterization of the complete signed graphs with a valid drawing in $\mathbb{R}$. Actually, the authors of [7] proved that a complete signed graph has a valid drawing in $\mathbb{R}$ if and only if its positive subgraph is a unit interval graph. The SCFE problem in the real line also was studied as an optimization problem by Pardo et al. in [18]. In that work, the authors defined as an error a violation of the inequality in Definition 2.1 and provided optimization algorithms that construct a drawing attempting to minimize the number of errors.

The SCFE problem also has been studied for different metric spaces. First, Benitez et al. in [2] studied the SCFE problem in the circumference. The authors of that work proved that the SCFE problem in the circumference is NP-Complete and gave a characterization of the complete signed graphs with a valid drawing. Indeed, they showed that a complete signed graph has a valid drawing in the circumference if and only if its positive subgraph is a proper circular arc graph. Later, Becerra in [1] studied the SCFE problem in trees. The main result of her work was to prove that a complete signed graph G has a valid drawing in a tree if and only if its positive subgraph is strongly chordal.

Spaen et al. in [23] studied the SCFE problem from a different perspective. They studied the problem of finding $L(n)$, the smallest dimension k such that any signed graph on n vertices has a valid drawing in $\mathbb{R}^k$, with respect to the Euclidean distance. They showed that $\log_5(n-3) \leq L(n) \leq n-2$.

Our Contributions. Our first contribution is to show that the Seriation and the SCFE problems are not the same. In Lemma 4.1, we show that seriation is a necessary condition for a valid drawing. Nevertheless, in Lemma 4.2, we show that seriation is

[1] It is worth noting that this value r denotes almost the same value as the r defined in the previous section. Actually, a zero entry in the matrix in the position ij denotes the absence of the edge $\{i, j\}$. Nevertheless, since the matrix is symmetric, if the ij entry is zero then the ji entry is also zero. Therefore, the r in this case counts twice a missing edge. However, that factor 2 does not change the complexity of the algorithm. Therefore, for simplicity we chose to abuse the notation.

not sufficient for a valid drawing.

The weighted version versus the signed original version of the SCFE problem does not allow a characterization of the set of graphs with a valid drawing in $\mathbb{R}$ via a subgraph of them, as it was done in previous works. Instead, for each weighted graph G, we define a polyhedron $M(G)\mathbf{x} \leq \mathbf{b}$ to provide a characterization of the set of weighted graphs with a valid drawing in $\mathbb{R}$. Indeed, we show in Theorem 4.4 that a weighted graph G has a valid drawing in $\mathbb{R}$ if and only if its polyhedron $M(G)\mathbf{x} \leq \mathbf{b}$ is not empty.

Our first result applied to complete weighted graphs allows us to conclude in Corollary 4.5 that given a complete weighted graph G, determining whether G has a valid drawing in $\mathbb{R}$, and finding one if applicable, can be done in polynomial time.

On the other hand, when the weighted graph is not complete, the previous result does not apply anymore. As we show Theorem 5.1, recognition of incomplete Robinsonian matrices is NP-complete, therefore, the construction of the polyhedron $M(G)\mathbf{x} \leq \mathbf{b}$ cannot be done in polynomial time (unless P=NP).

Nevertheless, we show in Theorem 5.4 that recognition of incomplete Robinsonian matrices can be done in time $O(n^2 \cdot L^r)$, where r is the number of zero entries in the matrix, and L is the number of different values in the matrix. Hence, in Corollary 5.5 we show that if the value r is a constant, determining whether an incomplete weighted graph G has a valid drawing in $\mathbb{R}$ can be done in polynomial time.

4. The Weighted SCFE Problem in the line. We start this section by showing that having a Robinsonian similarity matrix is a necessary condition to have a valid drawing in $\mathbb{R}$.

LEMMA 4.1. *Let G be a weighted graph. If G has a valid drawing in $\mathbb{R}$, $A(G)$ is Robinsonian.*

Proof. Let $G = (V, E)$ be a weighted graph with weight function w. Let $D : V \to \mathbb{R}$ be a valid drawing of G in $\mathbb{R}$. The valid drawing D determines an ordering on the set of vertices V. Indeed, for i and j in V, we say that $i <_D j$ if $D(i) < D(j)$. We show that if $A(G)$ is written using the ordering determined by D for its rows and columns, it will be in Robinsonian form.

Consider any i, j and k such that $i < j < k$ and $A(G)_{ik} \neq *$, $A(G)_{ij} \neq *$ and $A(G)_{jk} \neq *$. Since D is a valid drawing and $d(D(i), D(k)) > d(D(i), D(j))$, then $A(G)_{ik} \leq A(G)_{ij}$. Equivalently, since D is a valid drawing and $d(D(i), D(k)) > d(D(j), D(k))$, then, $A(G)_{ik} \leq A(G)_{jk}$. Therefore, $A(G)_{ik} \leq \min\{A(G)_{ij}, A(G)_{jk}\}$.

In conclusion, $A(G)$, the similarity matrix of G, is Robinsonian, and when it is written according to the ordering determined by any valid drawing of G in $\mathbb{R}$ it is in Robinsonian form. ∎

Nevertheless, having a Robinsonian similarity matrix is not enough.

LEMMA 4.2. *There exists a weighted graph G with Robinsonian similarity matrix, but, without a valid drawing in $\mathbb{R}$.*

Proof. Let G be the complete weighted graph with vertex set $\{a, b, c, d, e\}$ and similarity matrix

$$A(G) = \begin{bmatrix} 5 & 2 & 2 & 1 & 1 \\ 2 & 5 & 3 & 2 & 1 \\ 2 & 3 & 5 & 4 & 1 \\ 1 & 2 & 4 & 5 & 5 \\ 1 & 1 & 1 & 5 & 5 \end{bmatrix}$$

written with rows and columns ordered as a, b, c, d, e. $A(G)$ is Robinsonian, nevertheless, we will show by contradiction that G does not have a valid drawing in $\mathbb{R}$.

Assume that G has a valid drawing D in $\mathbb{R}$. Since the order a, b, c, d, e of the rows and columns of $A(G)$ is the only one that presents $A(G)$ in Robinsonian form, then D has to be such that

$$D(a) < D(b) < D(c) < D(d) < D(e). \tag{4.1}$$

Since D is a valid drawing, the following inequalities hold:

$$D(b) - D(a) > D(c) - D(b) \tag{4.2}$$
$$D(e) - D(b) > D(b) - D(a) \tag{4.3}$$
$$D(c) - D(b) > D(d) - D(c) \tag{4.4}$$
$$D(e) - D(c) > D(c) - D(a) \tag{4.5}$$
$$D(d) - D(c) > D(e) - D(d). \tag{4.6}$$

Without loss of generality, assume that $D(a) = 0$. Then, from inequalities (4.1) and (4.2) we obtain:

$$D(b) < D(c) < 2D(b). \tag{4.7}$$

On the other hand, from inequalities (4.5) and (4.6), we obtain $2D(c) < D(e) < 2D(d) - D(c)$, which implies:

$$3D(c) < 2D(d). \tag{4.8}$$

Finally, inequality (4.4) is equivalent to $2D(d) < 4D(c) - 2D(b)$, which, together with (4.8), implies $2D(b) < D(c)$. But, the last inequality contradicts inequality (4.7). ∎

The goal of the rest of this section is to transform the weighted SCFE problem in the real line into the problem of finding a point in a convex polyhedron. Actually, given a weighted graph G, we define a convex polyhedron $M(G)\mathbf{x} \leq \mathbf{b}$, where each point $\mathbf{x} = (x_1, x_2, \ldots, x_n)$ in the convex polyhedron is a valid drawing of G in $\mathbb{R}$. Indeed, for any given $\mathbf{x}$ in $M(G)\mathbf{x} \leq \mathbf{b}$, each variable x_i represents the position of vertex i in the real line for that valid drawing. Therefore, finding a point in $M(G)\mathbf{x} \leq \mathbf{b}$ is equivalent to find a valid drawing for G in $\mathbb{R}$.

We first remark that if a given weighted graph G has a valid drawing in $\mathbb{R}$, it actually has an infinite number of them. Indeed, given a valid drawing in $\mathbb{R}$ for a weighted graph G, one can obtain a different valid drawing for the same graph by summing or multiplying each vertex position by any positive constant. The second case (when each position is multiplied by a positive constant) is important for us, because it allows to state the following lemma.

LEMMA 4.3. *Let G be a weighted graph with a valid drawing in $\mathbb{R}$. Then, for any $\epsilon > 0$ there exists a valid drawing D_ϵ of G in $\mathbb{R}$ such that:*

$$\min_{1 \leq i < n} D_\epsilon(i+1) - D_\epsilon(i) \geq \epsilon.$$

Proof. Let G be a weighted graph with a valid drawing D in $\mathbb{R}$. We consider without loss of generality that $1 <_D 2 <_D 3 <_d \ldots <_D n$. Consider any $\epsilon > 0$. Let $\delta = \min_{1 \leq i < n} D(i+1) - D(i)$ be the minimum distance between two consecutive vertices in the drawing. Multiply every $D(i)$ by ϵ/δ. Therefore, we obtain a new valid drawing D_ϵ defined as $D_\epsilon(i) = \epsilon D(i)/\delta$, such that $\min_{1 \leq i < n} D_\epsilon(i+1) - D_\epsilon(i) = \epsilon$. ∎

Now, we proceed with the construction of the matrix $M(G)$ and the vector $\mathbf{b}$ of the convex polyhedron $M(G)\mathbf{x} \leq \mathbf{b}$. By Lemma 4.1, the ordering defined by a valid drawing makes $A(G)$ to be in its Robinsonian form. Assume that G is a weighted graph with Robinsonian similarity matrix. Moreover, consider $A(G)$ to be in Robinsonian form. Therefore, if we want to construct a valid drawing D in $\mathbb{R}$ for G, the vertices should be ordered in the same way as the rows and columns of $A(G)$. Hence, if the *i-th* row (or column) of $A(G)$ contains the similarities of vertex i, then $D(1) < D(2) < \cdots < D(n)$. Therefore, we want $x_1 < x_2 < \cdots < x_n$. Now, considering Lemma 4.3, we write the following set of restrictions for any $\epsilon > 0$:

$$x_i - x_{i+1} \leq -\epsilon, \quad \forall i \in \{1, 2, 3, \ldots, n-1\}. \tag{4.9}$$

This restrictions are called *ordering restrictions*.

On the other hand, each row of $A(G)$ provides two types of restrictions. We call these restrictions *right with respect to left* and *left with respect to right* restrictions. Right with respect to left restrictions are obtained as follows. For each row j and for every index $k > j$, let $i(k)$ be the largest index such that $i(k) < j$ and $A(G)_{ji(k)} < A(G)_{jk}$. Therefore, since $A(G)_{ji(k)} < A(G)_{jk}$, vertices j and k are more similar between them than vertices j and $i(k)$. Hence, in any valid drawing D it must occur $D(k) - D(j) < D(j) - D(i(k))$. We transform this strict inequality into the following restriction for a sufficiently small $\epsilon > 0$:

$$x_{i(k)} - 2x_j + x_k \leq -\epsilon, \quad \forall j \in \{2, 3, \ldots, n-1\} \text{ and } \forall k > j. \tag{4.10}$$

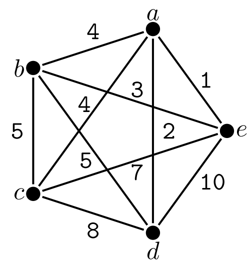
Left with respect to right restrictions are symmetrical to the previous restriction. For each row j and for every index $i < j$, let $k(i)$ be the smallest index such that $j < k(i)$ and $A(G)_{ji} > A(G)_{jk(i)}$. Therefore, since $A(G)_{ji} > A(G)_{jk(i)}$, vertices i and j are more similar between them than vertices j and $k(i)$. Hence, in any valid drawing D, it must occur $D(j) - D(i) < D(k(i)) - D(j)$. We transform this strict inequality into the following restriction for a sufficiently small $\epsilon > 0$:

$$-x_i + 2x_j - x_{k(i)} \leq -\epsilon, \quad \forall j \in \{2, 3, \ldots, n-1\} \text{ and } \forall i < j. \tag{4.11}$$

It is worth mentioning that some of the inequalities described in equations (4.10) and (4.11) may be obtained from inequalities presented in Equation (4.9) and different inequalities described in equations (4.10) and (4.11). Hence, some restrictions may be redundant. In an attempt to keep the presentation of this document clean and simple, we omit a discussion in this regard. It is worth mentioning though that it does not impact the results of this document.

Given a weighted graph G with n vertices, the *matrix of restrictions* of G (or the *matrix of coefficients* of G), denoted by $M(G)$, is the matrix that includes the $n-1$ ordering restrictions, the at most $(n-2)(n-1)/2$ right with respect to left restrictions, and the at most $(n-2)(n-1)/2$ left with respect to right restrictions. In total, the matrix $M(G)$ has $h \leq (n-1)^2$ rows and n columns. On the other hand, the vector $\mathbf{b}$ is a $h \times 1$ vector with a $-\epsilon$ in every entry. An example of a weighted graph, its similarity matrix in Robinsonian form, and its corresponding matrix of restrictions is given in Figure 1.

Now, we want to show that for any weighted graph G with Robinsonian similarity matrix, the convex polyhedron $M(G)\mathbf{x} \leq \mathbf{b}$ is not empty if and only if G has a valid drawing in $\mathbb{R}$.



(a)

$$
\begin{array}{c|ccccc}
 & a & b & c & d & e \\
\hline
a & 10 & 4 & 4 & 2 & 1 \\
b & 4 & 10 & 5 & 5 & 3 \\
c & 4 & 5 & 10 & 8 & 7 \\
d & 2 & 5 & 8 & 10 & 10 \\
e & 1 & 3 & 7 & 10 & 10
\end{array}
$$

(b)

$$
\begin{bmatrix}
1 & -1 & 0 & 0 & 0 \\
0 & 1 & -1 & 0 & 0 \\
0 & 0 & 1 & -1 & 0 \\
0 & 0 & 0 & 1 & -1 \\
1 & -2 & 0 & 1 & 0 \\
-1 & 2 & 0 & 0 & -1 \\
0 & 1 & -2 & 0 & 1 \\
0 & 0 & 1 & -2 & 1
\end{bmatrix}
$$

(c)

FIG. 1. *Example of a complete weighted graph, its similarity matrix, and its corresponding matrix of restrictions. Subfigure (a) shows a complete weighted graph. Subfigure (b) shows its similarity matrix written in its Robinsonian form. It also shows the order of the vertices in which the similarity matrix is written. Subfigure (c) shows the restriction matrix for the weighted graph in Subfigure (a). In the first 4 rows appear the ordering restrictions. Rows five and six show the right with respect to left and left with respect to right restrictions for vertex b. Rows seven and eight show right with respect to left restrictions for vertices c and d, respectively.*

THEOREM 4.4. *Let G be a weighted graph with Robinsonian similarity matrix. Let $M(G)$ be the $h \times n$ matrix of restrictions of G. Let b be the $h \times 1$ vector with $-\epsilon < 0$ in every entry. Then, G has a valid drawing in $\mathbb{R}$ if and only if the polyhedron $M(G)\mathbf{x} \leq \mathbf{b}$ is not empty.*

Proof. Let G be a weighted graph with valid drawing in $\mathbb{R}$. Let D be a valid drawing of G in $\mathbb{R}$. Label the vertices of G according to the order implied by D, i. e., the left most vertex in D is vertex 1, the next vertex is vertex 2 and so on until vertex n. By construction of $M(G)\mathbf{x} \leq \mathbf{b}$, for any $\epsilon > 0$, D can be scaled to a valid drawing D' such that the vector $(D'(1), D'(2), \ldots, D'(n))$ belongs to the polyhedron $M(G)\mathbf{x} \leq \mathbf{b}$.

On the other hand, assume that the polyhedron $M(G)\mathbf{x} \leq \mathbf{b}$ is not empty. Let $x = (x_1, x_2 \ldots, x_n)$ be a point in $M(G)\mathbf{x} \leq \mathbf{b}$. Label the vertices of G according to the columns of its similarity matrix written in Robinsonian form, i. e., the vertex i is the vertex corresponding to the *i-th* column of $A(G)$. Now, consider the drawing D of G in $\mathbb{R}$ defined as follows: $D(i) = x_i$ for all $1 \leq i \leq n$.

We show now that D is a valid drawing. Assume that D is not a valid drawing. Therefore, there exist three vertices i, j and k such that $A_{ij} < A_{ik}$, but $|D(i)-D(j)| \leq |D(i) - D(k)|$. Note that the last inequality is not valid if $D(i) < D(k) < D(j)$ or if $D(j) < D(k) < D(i)$, therefore, these cases are discarded. If $D(i) < D(k) < D(j)$ or $D(j) < D(k) < D(i)$, there is a contradiction since $A_{ij} < A_{ik}$, and, in that case, $A(G)$ would not be in Robinsonian form.

Assume that $D(j) < D(i) < D(k)$. Therefore, $|D(i) - D(j)| \leq |D(i) - D(k)|$ becomes $D(i) - D(j) \leq D(k) - D(i)$, or equivalently, $0 \leq D(j) - 2D(i) + D(k)$. Nevertheless, since $A_{ij} < A_{ik}$, the right with respect to left restriction $x_j - 2x_i + x_k \leq -\epsilon$ is included in $M(G)\mathbf{x} \leq \mathbf{b}$. Therefore, since D comes from a point in $M(G)\mathbf{x} \leq \mathbf{b}$, $D(j) - 2D(i) + D(k) \leq -\epsilon$, which is a contradiction since $\epsilon > 0$.

If we assume now $D(k) < D(i) < D(j)$, then $|D(i) - D(j)| \leq |D(i) - D(k)|$ becomes $0 \leq -D(k) + 2D(i) - D(j)$. Nevertheless, since $A_{ij} < A_{ik}$, the left with respect to right restriction $-x_k + 2x_i - x_j \leq -\epsilon$ is included in $M(G)\mathbf{x} \leq \mathbf{b}$. By equivalent arguments than before, we achieve a contradiction.

Therefore, the condition $|D(i)-D(j)| \leq |D(i)-D(k)|$ is not possible, and hence, D is a valid drawing. ∎

The weighted SCFE problem now is equivalent to find a point in a convex polyhedron. If the valid drawings are restricted to be nonnegative, then the SCFE problem can be treated as a linear program. Because, if the polyhedron $M(G)\mathbf{x} \leq \mathbf{b}$ is not empty, there is always a point $\mathbf{x}$ in $M(G)\mathbf{x} \leq \mathbf{b}$ with $x_0 = 0$. Therefore, the SCFE problem is equivalent to find $\min x_0$ subject to $M(G)\mathbf{x} \leq \mathbf{b}$, and nonnegative $\mathbf{x}$.

On the other hand, it is required to have $A(G)$ in Robinsonian form to construct $M(G)$. Since complete Robinsonian matrices can be recognized in time $O(n^2)$, it is possible to construct the matrix $M(G)$ in polynomial time when G is complete. Therefore, we can state the following corollary.

COROLLARY 4.5. *Let G be a complete weighted graph. Deciding whether G has a valid drawing in $\mathbb{R}$ can be done in polynomial time. Moreover, a valid drawing for G in $\mathbb{R}$ can be computed also in polynomial time if such drawing exists.*

5. The Weighted SCFE Problem for Incomplete Weighted Graphs.

If the condition of being complete is not requested for the weighted graph, it is not possible to determine in polynomial time whether its similarity matrix is Robinsonian or not, unless P=NP. Indeed, we now show that Robinsonian matrix recognition in the general case is NP-Complete.

THEOREM 5.1. *The Robinsonian matrix recognition problem in the general case is NP-Complete.*

Proof. In order to prove the Theorem, we reduce the graph sandwich problem for unit interval graphs to the Robinsonian matrix recognition problem.

The graph sandwich problem for unit interval graphs is the problem of finding a unit interval graph that is *sandwiched* between two other graphs, one of which must be a subgraph and the other of which must be a supergraph of the desired graph. Indeed, an instance of the graph sandwich problem for unit interval graphs is a vertex set V, a mandatory edge set E^1, and a larger edge set E^2, such that $E^1 \subseteq E^2 \subseteq V \times V$. The question then is to decide the existence of a graph $G = (V, E)$ such that $E^1 \subseteq E \subseteq E^2$ and G is a unit interval graph.

From an instance of the graph sandwich problem for unit interval graphs, we construct an instance for the Robinsonian matrix recognition problem as follows. Let A be the symmetric matrix defined as:

$$A_{ij} = \begin{cases} 1 & \text{if } i = j, \\ 1 & \text{if } \{i,j\} \in E^1, \\ * & \text{if } \{i,j\} \in E^2 \setminus E^1, \\ 0 & \text{if } \{i,j\} \notin E^2. \end{cases}$$

The relationship between Robinsonian matrices and unit interval graphs presented in [21] says that, A is Robinsonian if and only if there exists a unit interval graph $G = (V, E)$ such that $E^1 \subseteq E \subseteq E^2$.

Furthermore, since the graph sandwich problem for unit interval graphs is NP-Complete [10], and Robinsonian matrix recognition in the general case belongs to NP, we can say that Robinsonian matrix recognition is NP-Complete. ∎

Let A be an incomplete similarity matrix. Every pair $\{i, j\}$ such that $A_{ij} = *$ is a missing entry of A. A *completion* of A is an assignment of values to all the missing entries of A. We say that *a completion of A is Robinsonian* if and only if the completed matrix is Robinsonian. Let $S \subseteq \mathbb{R}$ be a set of real values. A completion of A whose new values are taken from S is said to be a *completion of A with values in S*. Let $p \in \mathbb{R}$ be any real value, we define $\lceil p \rceil_S := \min_{s \in S}\{s : s \geq p\}$ and $\lfloor p \rfloor_S := \max_{s \in S}\{s : s \leq p\}$. We define the set of entry values of A as the set $w(A) := \{A_{ij} \in \mathbb{R}\}$. Now, we state the following lemma for incomplete similarity matrices.

LEMMA 5.2. *Let G be an incomplete weighted graph and A be its incomplete similarity matrix. A is Robinsonian if and only if A has a Robinsonian completion.*

Proof. Let G be an incomplete weighted graph and A be its incomplete similarity matrix. If A has a Robinsonian completion, then one can write this completion of A in Robinsonian form and delete all the added entries. The outcome is A written in Robinsonian form.

On the other hand, if A is Robinsonian, we can write it in Robinsonian form and complete it as follows. For every missing entry A_{ij} with $1 \leq i < j \leq n$ define $A_{ij} = \min A_{ij-1}, A_{i+1j}$. Since none entry of the diagonal is missing, this completion always can be done moving away from the diagonal. Finally, by construction the completion is Robinsonian. ◻

LEMMA 5.3. *Let G be an incomplete weighted graph and A be its incomplete similarity matrix with set of entry values $w(A)$. A has a Robinsonian completion with values in $\mathbb{R}$ if and only if A has a Robinsonian completion with values in $w(A)$.*

Proof. On one hand, since $w(A) \subseteq \mathbb{R}$, if A has a Robinsonian completion with values in $w(A)$, then it also has a Robinsonian completion with values in $\mathbb{R}$.

Now, assume that A has a Robinsonian completion A' with values in $\mathbb{R}$. Assume that A' is in Robinsonian form. We construct then a Robinsonian completion A'' from A' with values in $w(A)$ as follows:

$$A''_{ij} = \begin{cases} A'_{ij} & \text{if } A_{ij} \neq *, \\ \lfloor A'_{ij} \rfloor_{w(A)} & \text{if } A_{ij} = * \quad \wedge \quad A'_{ij} > A_{ts} \text{ for all } A_{ts} \in w(A), \\ \lceil A'_{ij} \rceil_{w(A)} & \text{if } A_{ij} = * \quad \wedge \quad \exists A_{ts} \in w(A) \text{ such that } A_{ts} > A'_{ij}, \end{cases}$$

We finish the proof by showing that A'' is in Robinsonian form. Consider $1 \leq i < j \leq n$, we want to show that $A''_{ij} \leq \min\{A''_{ij-1}, A''_{i+1j}\}$. By contradiction, assume that $A''_{ij} > A''_{ij-1}$. Therefore, by construction, $A'_{ij} > A'_{ij-1}$. Equivalently, $A''_{ij} > A''_{i+1j}$ implies that $A'_{ij} > A'_{i+1j}$. In any case, any of these two conclusions creates a contradiction, since A' is Robinsonian and it is in Robinsonian form, therefore $A'_{ij} \leq \min\{A'_{ij-1}, A'_{i+1j}\}$. ◻

As a consequence of the previous lemma, we state the following theorem.

THEOREM 5.4. *Let G be a weighted graph with r missing edges and L different value weights. Then, it is possible to decide if $A(G)$ is Robinsonian in time $O(n^2 \cdot L^r)$.*

Proof. Let G be a weighted graph with r missing edges and L different value weights. Let $A(G)$ be its similarity matrix. There exist L^r different completions of $A(G)$ with values in $w(A(G))$. Therefore, an exhaustive search over all the completions of $A(G)$ with values in $w(A(G))$ and testing for each of them the Robinsonian property takes $O(n^2 \cdot L^r)$. ◻

COROLLARY 5.5. *The weighted SCFE problem for an incomplete weighted graph G with r missing edges, where r is a constant that does not depend on n, can be solved in polynomial time.*

6. Final Remarks. Interestingly, in this work we show that the Seriation and the SCFE problems are not the same. Nevertheless, there are cases in which they are equivalent. For instance, an exhaustive analysis shows that if a weighted graph has at most four vertices then its similarity matrix is Robinsonian if and only if it has a valid drawing in $\mathbb{R}$. Whereas, in the proof of Lemma 4.2 we present a weighted graph with five vertices where seriation is not sufficient.

The Seriation and the SCFE problems are also equivalent if the number of different weights is not too big. The results presented in [21] and in [7], allow us to conclude that when there are two different weights then having a Robinsonian similarity matrix is equivalent to have a valid drawing in $\mathbb{R}$. Nevertheless, in the proof of Lemma 4.2 we show an example of a weighted graph with five different weights where where seriation is not enough. This final remark rises an interesting question, when this separation between the Seriation and the SCFE problem occurs?. Is the Seriation problem equivalent to the SCFE problem when the graph has four different weights?.

REFERENCES

[1] R. Becerra, *Caracterización y reconocimiento de grafos signados con dibujo válido en un árbol*, master's thesis, Departamento de Ingeniería Matemática, Universidad de Concepción, 2018.

[2] F. Benítez, J. Aracena, and C. Thraves Caro, *The sitting closer to friends than enemies problem in the circumference*, arXiv preprint arXiv:1811.02699, (2018).

[3] M. J. Brusco and S. Stahl, *Branch-and-Bound applications in combinatorial data analysis*, Springer Science & Business Media, 2006.

[4] V. Chepoi and B. Fichet, *Recognition of Robinsonian dissimilarities*, Journal of Classification, 14 (1997), pp. 311–325.

[5] V. Chepoi, B. Fichet, and M. Seston, *Seriation in the presence of errors: NP-hardness of l_∞-fitting Robinson structures to dissimilarity matrices*, Journal of classification, 26 (2009), pp. 279–296.

[6] V. Chepoi and M. Seston, *Seriation in the presence of errors: A factor 16 approximation algorithm for l_∞-fitting Robinson structures to distances*, Algorithmica, 59 (2011), pp. 521–568.

[7] M. Cygan, M. Pilipczuk, M. Pilipczuk, and J. O. Wojtaszczyk, *Sitting closer to friends than enemies, revisited*, Theory of Computing Systems, 56 (2015), pp. 394–405.

[8] C. Ding and X. He, *Linearized cluster assignment via spectral ordering*, in Proceedings of the twenty-first international conference on Machine learning, ACM, 2004, p. 30.

[9] D. Fortin, *Robinsonian matrices: Recognition challenges*, Journal of Classification, 34 (2017), pp. 191–222.

[10] M. C. Golumbic, H. Kaplan, and R. Shamir, *Graph sandwich problems*, Journal of Algorithms, 19 (1995), pp. 449–473.

[11] L. Hubert, P. Arabie, and J. Meulman, *Combinatorial data analysis: Optimization by dynamic programming*, vol. 6, SIAM, 2001.

[12] A.-M. Kermarrec and C. Thraves Caro, *Can everybody sit closer to their friends than their enemies?*, in Proceedings of the 36th International Symposium on Mathematical Foundations of Computer Science, Springer, 2011, pp. 388–399.

[13] M. Laurent and M. Seminaroti, *A Lex-BFS-based recognition algorithm for Robinsonian matrices*, Discrete Applied Mathematics, 222 (2017), pp. 151–165.

[14] M. Laurent and M. Seminaroti, *Similarity-first search: A new algorithm with application to Robinsonian matrix recognition*, SIAM Journal on Discrete Mathematics, 31 (2017), pp. 1765–1800.

[15] M. Laurent, M. Seminaroti, and S.-i. Tanigawa, *A structural characterization for certifying Robinsonian matrices*, arXiv preprint arXiv:1701.00806, (2017).

[16] I. Liiv, *Seriation and matrix reordering methods: An historical overview*, Statistical Analysis and Data Mining: The ASA Data Science Journal, 3 (2010), pp. 70–91.

[17] B. G. Mirkin and S. N. Rodin, *Graphs and Genes*, Springer-Verlag, 1984.

[18] E. G. Pardo, M. Soto, and C. Thraves Caro, *Embedding signed graphs in the line*, Journal of Combinatorial Optimization, 29 (2015), pp. 451–471.

[19] W. M. F. Petrie, *Sequences in prehistoric remains*, Journal of the Anthropological Institute

of Great Britain and Ireland, (1899), pp. 295–301.

[20] P. Préa and D. Fortin, *An optimal algorithm to recognize Robinsonian dissimilarities*, Journal of Classification, 31 (2014), pp. 351–385.

[21] F. S. Roberts, *Indifference graphs*, in Proof Techniques in Graph Theory, F. Harary, ed., Academic Press, New York, 1969, pp. 139–146.

[22] W. S. Robinson, *A method for chronologically ordering archaeological deposits*, American Antiquity, 16 (1951), pp. 293 – 301, https://doi.org/10.2307/276978.

[23] Q. Spaen, C. Thraves Caro, and M. Velednitsky, *The dimension of signed graph valid drawing*, Tech. Report 09, Departamento de Ingeniería Matemática, Universidad de Concepción, 2017.

[24] Y.-J. Tien, Y.-S. Lee, H.-M. Wu, and C.-H. Chen, *Methods for simultaneously identifying coherent local clusters with smooth global patterns in gene expression profiles*, BMC bioinformatics, 9 (2008), p. 155.