Consistency of the spectral seriation algorithm

Amine Natik

A thesis submitted in partial fulfillment of the requirements for the Master of Science degree in Mathematics and Statistics[1]

Department of Mathematics and Statistics
Faculty of Science
University of Ottawa

© Amine Natik, Ottawa, Canada, 2019

[1]The M.Sc. program is a joint program with Carleton University, administered by the Ottawa-Carleton Institute of Mathematics and Statistics

Abstract

Given n arbitrary objects $x_1, x_2, \ldots, x_n$ and a similarity matrix $P = (p_{i,j})_{1 \le i,j \le n}$, where $p_{i,j}$ measures the similarity between x_i and x_j. If the objects can be ordered along a linear chain so that the similarity decreases as the distance increase within this chain, then the goal of the seriation problem is to recover this ordering π given only the similarity matrix. When the data matrix P is completely accurate, the true relative order can be recovered from the spectral seriation algorithm [1]. In most applications, the matrix P is noisy, but the basic spectral seriation algorithm is still very popular. In this thesis, we study the consistency of this algorithm for a wide variety of statistical models, showing both consistency and bounds on the convergence rates. More specifically, we consider a model matrix P satisfying certain assumptions, and construct a noisy matrix $\widehat{P}$ where the input (i, j) is a coin flip with probability $p_{i,j}$. We show that the output $\hat{\pi}$ of the spectral seriation algorithm for the random matrix is very close to the true ordering π.

Dedications

I dedicate this thesis to my dear mother Saïda Hanine.

Acknowledgement

I would like to thank my supervisor Aaron Smith for his guidance and patience during my graduate studies. He is an important part of my success. I would also like to thank the professors, graduate students, and other employees of the Department of Mathematics and Statistics for their immense support, including Mayada El-Maalouf, Benoit Dionne, Diane Demers, Joseph Khoury, Mayer Alvo and Kelly Burkett. Many thanks to Pr. Iosif Pinelis who helped me, through the online forum *MathOverflow*, to prove an important inequality in my thesis. I would also like to thank "l'Académie Hassan II des Sciences et Techniques" for their generous financial support during my Master's. I would like Finally, I would like to thank my family and friends in Morocco and Canada for their encouragement throughout my educational career.

Contents

Chapter 1

Introduction

In many real-life problems, we are required to reconstruct the linear order between a set of variables given only their pairwise similarity. This task is important and has many applications especially in network analysis. For example, companies are interested in, based on users preferences, analyzing and ranking large datasets of music, movies, images, etc. For instance, suppose you read one thousand books and you want to rank them from the most preferred to the last preferred. It is usually hard to get an absolute ranking value to each book you read. However, if you only have two books, then it is easier to decide which of the two is better. Given the pairwise preferences, we can attempt to construct a ranking of the books (see e.g. [11]).

The seriation problem introduced by Robinson in [25] is a combinatorial problem where the goal is to order a set of items so that similar items are placed closer to each other. The pioneer of seriation is considered to be Flinders Petrie, an english egyptologist, who called it *sequence dating* in his paper [23]. While the seriation originates from archaeology, it has recent applications to ecology, sociology, biology, etc. In chapter 2 we will present a brief survey about the history of seriation.

The seriation problem translates mathematically to a problem of permutations over the rows and columns of symmetric matrices. Let n be a positive integer, from here on we denote by $\mathcal{P}_n$ the set of all permutations of the natural numbers $\{1, 2, \ldots, n\}$. We will often denote its elements by π while Π will refer to a permutation matrix, which is a binary $n \times n$ matrix defined by $\Pi_{ij} = 1$ if and only if $i = \pi(j)$. For a matrix $A \in \mathbb{R}^{n\times n}$ we denote by A^π the permutation of A by a permutation π, i.e., $A^\pi_{ij} = A_{\pi(i)\pi(j)}$ for all $1 \leq i, j \leq n$. Note that, $A^\pi = \Pi A \Pi^T$ where Π is the permutation matrix associated with π.

Given n elements $x_1, x_2, \ldots, x_n$ in a data stape $\mathcal{X}$, assume that we have a symmetric matrix A where for each $1 \leq i, j \leq n$, A_{ij} measures how similar x_i to x_j. We call A the *similarity matrix* of the x_i's. Furthermore, assume that the x_i's can be ordered along a chain, where the similarity between variables decreases with their distance within this chain. Then seriation's goal is to reconstruct this linear ordering given only the similarity matrix A as input. In metalanguage, the seriation problem seeks to find all the permutations $\pi \in \mathcal{P}_n$ such that if $i < j < k$ then $A^\pi_{i,j} \geq A^\pi_{i,k}$ and $A^\pi_{j,k} \geq A^\pi_{i,k}$, see Figure 1.1. In other words, we are trying to find the permutations π for which the elements $x_{\pi(1)}, x_{\pi(2)}, \ldots, x_{\pi(n)}$ form a linear

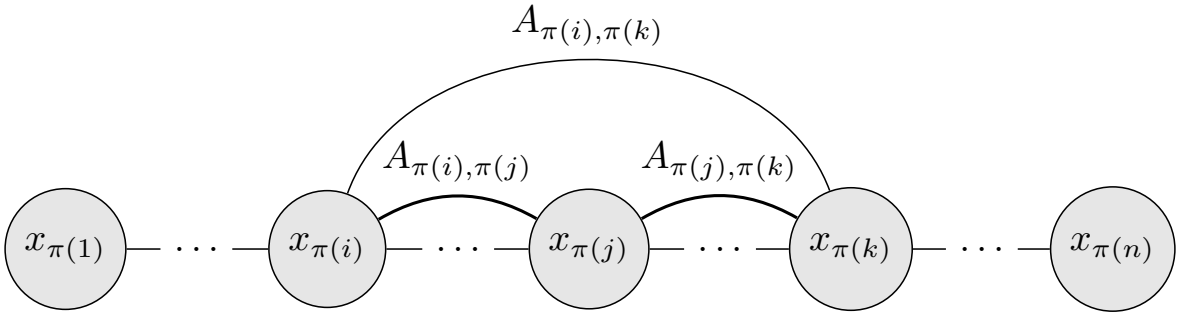


Figure 1.1: Linear ordering of elements $x_1, x_2, \ldots, x_n$. We can see that in this order the similarity decreases as the distance increase, in this figure $A_{\pi(i),\pi(k)} \leq A_{\pi(i),\pi(j)}$ and $A_{\pi(i),\pi(k)} \leq A_{\pi(j),\pi(k)}$

order, where the similarity decreases with the distance.

Given a $n \times n$ symmetric matrix A, we can see A as a similarity matrix of a set of n objects. If those objects form a linear ordering, where the similarity decrease with the distance, we say that A is a $\mathcal{R}$-matrix or *Robinson*-matrix. That is, if its entries are non-decreasing when moving toward the main diagonal, in each row or column. In other terms, for all $1 \leq i, j, k \leq n$ we have

$$i < j < k \Rightarrow A_{ij} \geq A_{ik} \text{ and } A_{jk} \geq A_{ik} \tag{1.0.1}$$

Figure 1.2 shows the shape of $\mathcal{R}$-matrices. Furthermore, we say that A is a pre-$\mathcal{R}$ matrix if there exists a permutation $\pi \in \mathcal{P}_n$ such that A^π is a $\mathcal{R}$-matrix. Figure 1.3 shows an example of a pre-$\mathcal{R}$ matrix on the left and the associated $\mathcal{R}$-matrix on the right.

Solving the seriation problem naively requires $n!$ steps, which is huge computationally for a large n. In chapter 3 we will present the spectral seriation algorithm 1 proposed in the paper [1] which solves the problem in an efficient way. The algorithm is based on reordering the matrix according to the components of the second eigenvector of the Laplacian of the matrix, called the Fiedler vector. According to [26], spectral methods were successfully applied in the past for relevant combinatorial problems such as recovering a partition, clique, coloring, bipartition, etc. Spectral methods are also used for the clustering problem (see e.g. [8]). Note that seriation is conceptually similar to clustering. However, while clustering aims to order the data into groups whose members are similar to each other, seriation seeks for a linear order of the objects such that similar objects are close to each other, which is more restrictive. Hence, although related, the two problems are substantially different.

The consecutive one problem or `C1P` for short is a problem that consists of permuting the rows of a binary matrix so as to make all the ones in each column appear consecutively. In particular [13] studied further the connection to interval graphs and proposed a graph algorithm to solve efficiently the `C1P`. David George Kendall in his paper [17] showed that a solution to `C1P` can be obtained by solving the seriation problem, in an other paper [18]

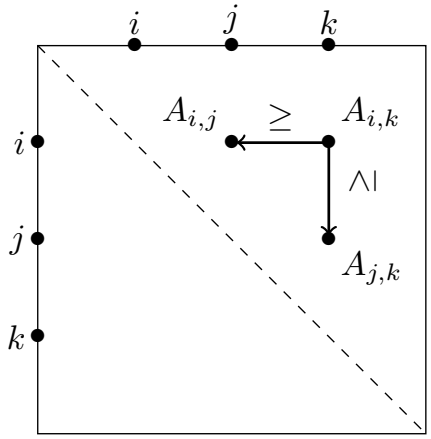
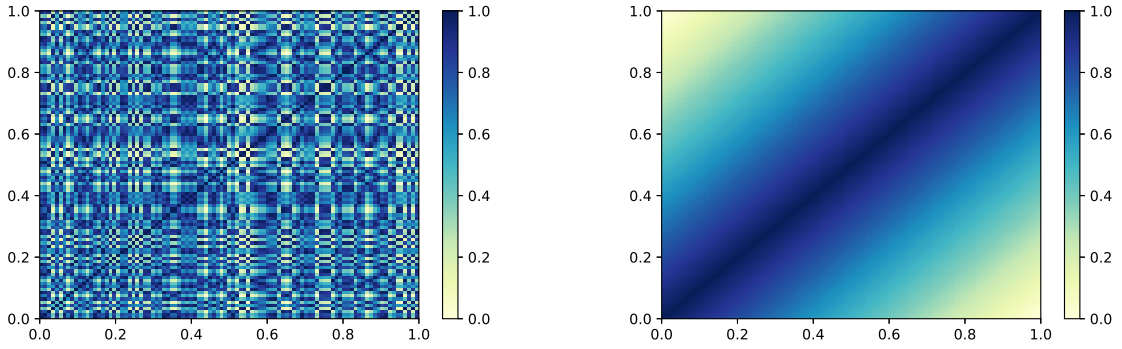


Figure 1.2: $\mathcal{R}$-matrices structure. The entries of the matrix are increasing as we move toward the main diagonal.

Kendall showed the relationship between the seriation problem and the problem of unimodal matrices, which is a generalization of the C1P. Other problems are also related to the seriation problem, for example in [19] the author present the relationship between the seriation problem and the travelling salesman problem. In [20] it was shown that the quadratic assignment problem (QAP) is easy for $\mathcal{R}$-matrices. The ranking problem is also related to the seriation problem as discussed previously, in [11] the authors considered the problem of ranking a set of n items given pairwise comparison between these items, and used the spectral seriation algorithm [1] to solve this task.

In practice it is natural to expect the data to be noisy due to missing information or incorrect labelling for example. In this case the reordering of the data-derived matrix may not be itself a $\mathcal{R}$-matrix but will be close to one. According to [16] the goal of seriation in the noisy settings is, given a matrix A, find a permutation π so that A^π is "almost a $\mathcal{R}$-matrix", for example we can choose π so that the quantity $\|A^\pi - R\|_p$ is minimized for all $\mathcal{R}$-matrices R, where $\|\cdot\|_p$ is the ℓ^p-norm for some p. Surprisingly despite decades of work on the seriation problem, little is known about the noisy seriation in the literature, that is, how robust and consistent to noise are the methods of seriation? In [10] a statistical approach using the least square estimate for the seriation problem is developed. In [16] the authors defined a new parameter Γ_1 which measures how badly a symmetric matrix A fails to be a $\mathcal{R}$-matrix, in particular $\Gamma_1(A) = 0$ if and only if A is a $\mathcal{R}$-matrix. Moreover, they proposed a polynomial time algorithm to solve the noisy seriation problem using the parameter Γ_1. In [12] the authors derived convex relaxations for the 2-SUM problem, which is a combinatorial problem which seeks to minimize a quadratic sum over permutations, in fact it is a special case of the quadratic assignment problem [20]. Moreover, it was showed in [12] that the 2-SUM is equivalent to the seriation problem in some settings. This approach improved the robustness to noise of the seriation problem.



(a) Original unordered matrix

(b) Ordered matrix

Figure 1.3: Example of a pre-$\mathcal{R}$ matrix on the left, and the latent $\mathcal{R}$-matrix on the right, after applying an appropriate permutation.

In this thesis we are interested in the consistency of the spectral seriation algorithm 1 when noise is present. More precisely, we are given an infinite family of pre-$\mathcal{R}$ matrices $(P^{(N)})_{N\in\mathbb{N}} \in [0,1]^{N\times N}$, we will call them the *model matrices*. Running the spectral seriation algorithm 1 on $P^{(N)}$ will return a permutation $\pi^{(N)}$ that will make $P^{(N)}$ a $\mathcal{R}$-matrix. Then, we construct a *random matrix* $\widehat{P}^{(N)}$ in the following way; for all $1 \leq i \leq j \leq N$ consider $\widehat{P}^{(N)}_{i,j} \sim \mathcal{B}\text{ernoulli}(P^{(N)}_{i,j})$, then put $\widehat{P}^{(N)}_{j,i} = \widehat{P}^{(N)}_{i,j}$ (since $\widehat{P}^{(N)}$ should be symmetric). Note that $\widehat{P}^{(N)}$ is not a perfect pre-$\mathcal{R}$ matrix, due to its random nature, but is very close to one. Running the the spectral seriation algorithm 1 will return a permutation $\widehat{\pi}^{(N)}$. The question we are concerned with is how close the true permutation $\pi^{(N)}$ to the random permutation $\widehat{\pi}^{(N)}$? Since the spectral seriation algorithm 1 requires the computation of the spectrum of the matrix, this question is very hard to answer in general, as far as we know, only two models where considered in the literature, in [26] the authors considered the same problem for the family of binary matrices $P^{(N)}$ defined by $P^{(N)}_{i,j} = 1$ if and only if $|i-j| \leq N/2$. In [11] the same question was answered for the model $P^{(N)} = 1 - |i-j|/N$ for all $1 \leq i,j \leq N$. Here we will answer the question of noisy seriation for a large family of model matrices. While in [11, 26] the authors computed the eigenvalues and eigenvectors to answer the question, here we will look at the limiting object of the matrices as a graphon [22], we show that under mild conditions on the graphon, the spectral seriation algorithm 1 is robust to noise. This same idea was used in [28] to show the consistency of the spectral clustering algorithm and in [2] to find the central node in a random network. Note that, graphon theory was also used in the contex of seriation in the paper [4] where the authors defined a new parameter Γ^* over graphs which aims to measure the similarity of the graph to an instance of a random graph with a linear embedding.

This thesis is organized as follows, in chapter 2 we will present an overview of the history of the seriation problem, in chapter 3 we will introduce the spectral seriation algorithm 1, and the relation between the seriation and the consecutive one problem, interval graphs

and the quadratic assignment problem. In chapter 4 we will present a brief overview of the noisy seriation in the literature. Finally, in chapter 5 we present our contribution to the consistency of the spectral seriation algorithm 1.

Chapter 2

Historical overview

2.1 Introduction

The seriation problem is an explanatory data analysis technique which consists of revealing the best ordering of a set of units in such a way that highly correlated pairs of elements are near each other. In this chapter we will give a brief historical overview about the seriation problem, for a detailed study we suggest [21, 14].

2.2 Seriation and Archaeology

The seriation problem has its roots in archaeology, as far as we know, the first systematic method for seriation was introduced and developed by an English Egyptologist W. M. Flinders Petrie [23] who called it *sequence dating*, he introduced seriation to sequence chronologically a set of graves, found in Egypt at the end of XIV century, based on the artifacts they share (see Figure 2.1). Since absolute dating techniques such as stratigraphy or radiocarbon dating were not available, Petrie invented a *relative dating* procedure to sequence the tombs. His approach was different from others for depending exclusively on the information and the similarity of artifacts versus professional human judgment of evolutionary and development complexity of artifacts. He assumed that the design styles of the potteries of the ancient Egyptian follow a bell curve of popularity – starting slowly, growing to a peak and then dying away as another style becomes popular (see Figure 2.3) – hence, he classified potteries in the tombs according to their style, and placed them in sequence so that graves with the most similar proportions of pottery styles are always closer in time.

Petrie listed the contents of each grave on a strip of cardboard (see Figure 2.2) and swapped the papers around until he arrived at a sequence he was satisfied with. He reasoned that the most accurate sequence would be the one where concentrations of certain design styles had the shortest duration across the sequence of papers. Whereas Petrie is considered the inventor of contextual seriation, Robinson [25] and Brainerd [3] were the first to address the problem of frequency seriation.

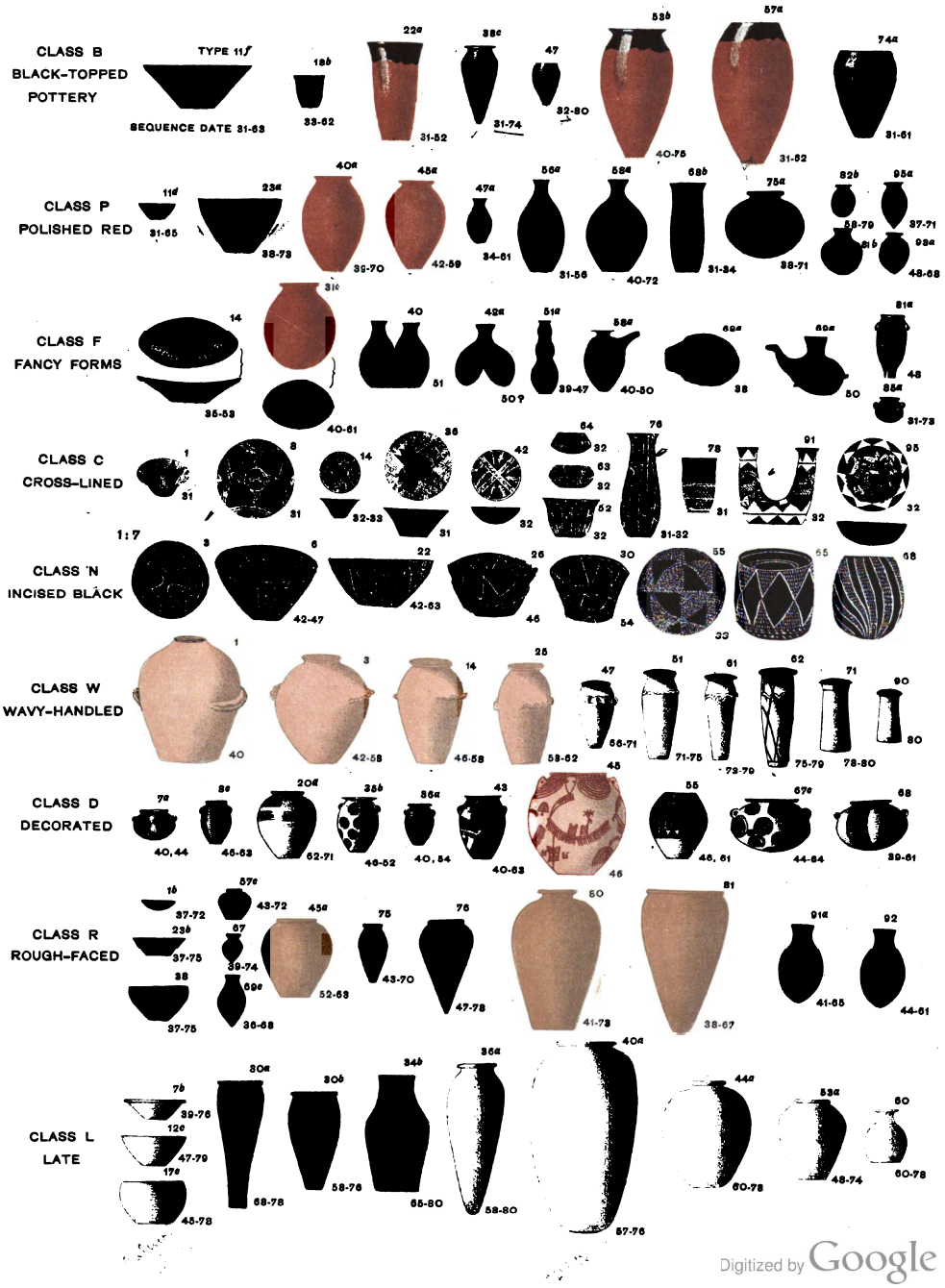
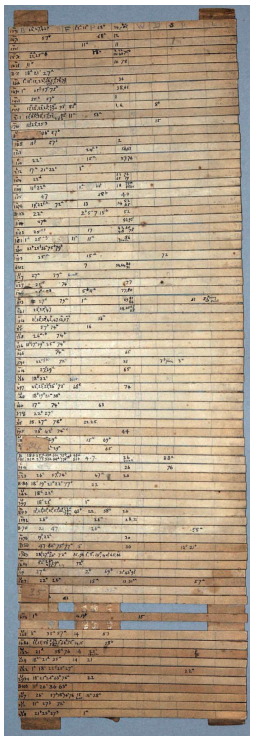
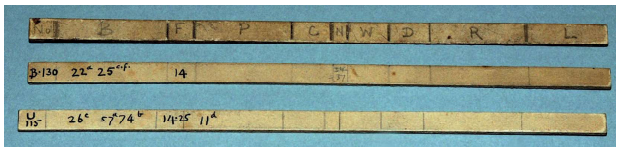


Figure 2.1: Types of pottery found by Petrie [14].



(a) Each row corresponds to a tomb, and each of the nine columns represent the pottery found according to his classification in Figure 2.1



(b) Petrie's arrangement of slips, partial view.

Figure 2.2: Paper-slips (© Courtesy of the Petrie Museum, London) [14]

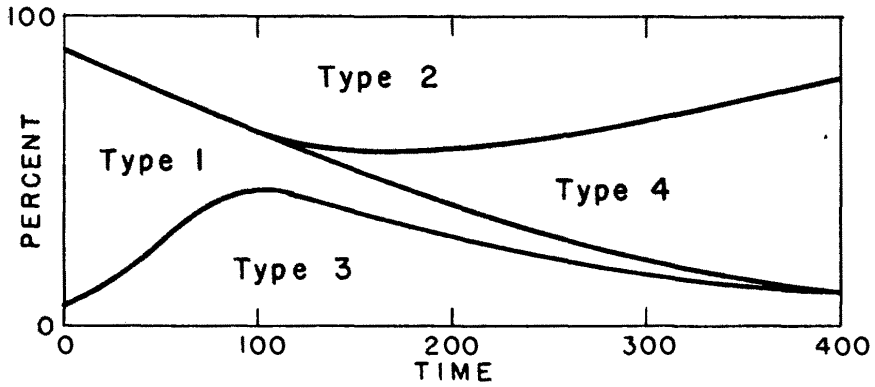


Figure 2.3: The differential use of four types of pottery over 400 years [25].

2.3 The first mathematical algorithm for seriation

Observations and methods presented by Petrie [23] were not written down using classical mathematical notations, but are nevertheless recognized for being the first to clearly formulate the idea of sequencing objects on the basis of their incidence or abundance. Till then, the evaluation of seriation results remained primarily intuitive and subjective until Robinson [25] proposed a mathematical algorithm for seriation. He was the first to propose a form of similarity matrix of the units, that will satisfy a 'perfect' chronological ordering: the highest values in the matrix should be along the diagonal and monotonically decrease when moving away from the diagonal. Nowadays, we refer to a matrix having this property as *Robinsonian matrix*, *Robinson matrix* or $\mathcal{R}$*-matrix*, in the next chapter we will discuss the mathematical properties of such matrices.

Robinson's study [25] is based upon the data from Mani collection of Brainerd's [3], Figure 2.4 shows the percentages of eight different pottery types appearing in three Mani trenches, they are labeled I, II, III. The deposits were stratigraphically collected, and the strata are denoted by A, B, and C – A representing the top or most recent stratum, B the next stratum underneath and C the bottom or earliest stratum. His method relies on the assumption that deposits which are close together in time will have similar percentage distributions, he defined a coefficient of agreement to measure the similarity distributions of the deposits. For example the similarity between deposits IIA and IIB is

$$\begin{aligned} 200 - \sum_{i=1}^{8} \left| p_i^{(\mathrm{IIA})} - p_i^{(\mathrm{IIB})} \right| &= 200 - |24.0 - 1.4| - \cdots - |3.9 - 0.0| \\ &= 4.6 \end{aligned}$$

After calculating the agreement coefficients for all the deposits, we get the rounded matrix in Figure 2.5a. The positive and negative signs in the table show whether an agreement coefficient of the table already had the desired pattern (increasing when moving toward the diagonal), the smaller the number of negative signs, the closer the approximation to the desired pattern. In Table 2.5a there are 60% of negative signs. We then rearrange the deposits so that their totals will show an ascending–descending pattern (bell curve), the result is shown in Table 2.5b, where there are only 38% of negative signs. The last step consists of encircling in each column half of those agreement coefficients having largest values, in order to exhibit the desired pattern we will rearrange the deposits so that those largest agreement coefficients are closer to the diagonal. This gives the Table 2.5c, where there is only 12% of negative signs, this is the best we can do. From that order, we can conclude that we have the following chronological order : IIA $<$ IIIA $<$ IIIB $<$ IA $<$ IIIC $<$ IB $<$ IIB $<$ IIC.

2.4 Conclusion

Seriation was first introduced and studied by archaeologists as a relative dating method, it was used to order chronologically different sites where absolute dating methods cannot be

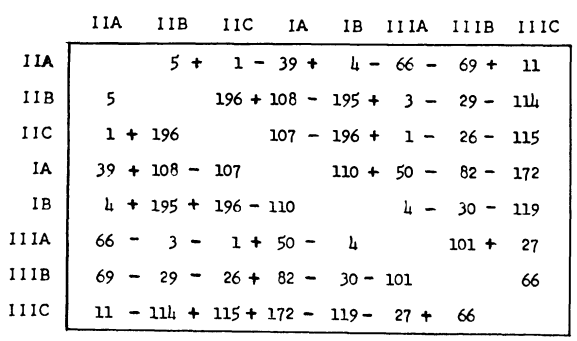
Type	IIA	IIB	IIC	IA	IB	IIIA	IIIB	IIIC
1	24.0	1.4	.2	11.3	.3	29.6	54.3	.0
2	66.8	.9	.0	.0	.0	.0	3.5	.0
3	1.3	.0	.2	3.8	.2	14.1	14.0	6.6
4	.0	.0	.0	1.3	.2	.0	1.8	3.3
5	.0	.0	.0	3.3	.5	.0	5.3	5.5
6	4.0	.0	.0	24.9	1.4	7.0	7.0	27.5
7	.0	97.7	99.3	52.6	97.4	.0	12.3	57.1
8	3.9	.0	.3	2.8	.0	49.3	1.8	.0
	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0

Figure 2.4: Percentages of eight types of pottery in three stratified trenches deposits [25].

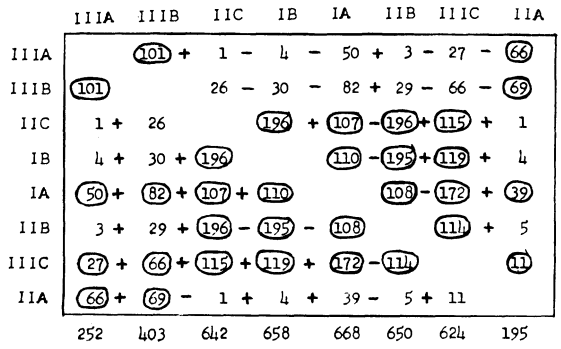
applied. Nowadays, seriation results are no longer produced manually in Petrie's slips but by appropriate algorithms. Though according to David George Kendall, Petrie's paper [23] showed already a deep understanding of the mathematics of the seriation problem. Quote:

"*..in my view Petrie should be ranked with the greatest applied mathematicians of the nineteenth century*"

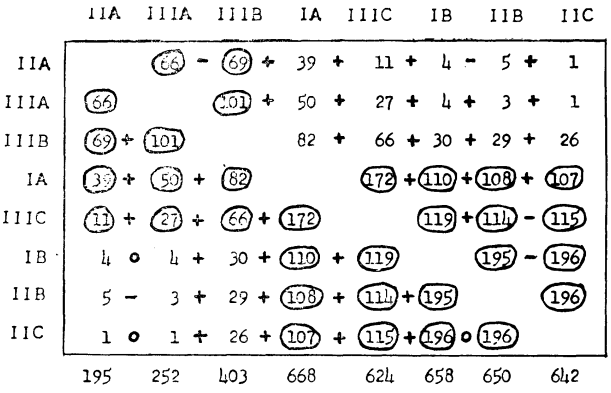
The paper of Robinson [25] is the first to give an algorithm for seriation, he based his frequency seriation method on a similarity matrix. About a decade later, several algorithms for chronological ordering were proposed. Seriation became popular and used in different disciplines including: Biology, Manufacturing, Anthropology, Sociology, Cartography.... In the next chapter we will study seriation from a mathematical point of view.



(a) 1st order.



(b) 2nd order.



(c) 3rd order.

Figure 2.5: Agreement coefficients for three stratified trenches, steps of the method in [25]

Chapter 3

The seriation problem

In this chapter we will present an overview of the seriation problem in the literature. In the first section, we present the main algorithm of this thesis, the spectral seriation algorithm 1. Then we discuss other related problems such as: the consecutive ones problem, the recognition of interval graphs, the quadratic assignment problem.

3.1 The spectral seriation algorithm

In this section we present the spectral seriation algorithm introduced by Atkins et al. in [1]. First, we start by introducing some definitions and notations. Let n be a positive integer, $\mathcal{P}_n$ denote the set of all permutations of the natural numbers $\{1, 2, \ldots, n\}$. For a vector $x = [x_1, \ldots, x_n]^T$ we denote x^π the permutation of x by π, i.e., $x^\pi = [x_{\pi(1)}, \ldots, x_{\pi(n)}]^T$. Similarly for a matrix $A \in \mathbb{R}^{n\times n}$ we denote by A^π the permutation of A by a permutation π (see the introduction chapter). A vector $x \in \mathbb{R}^n$ is said to be *monotone* if $x_1 \leq x_2 \leq \cdots \leq x_n$ or $x_1 \geq x_2 \geq \cdots \geq x_n$. Denote by $e \in \mathbb{R}^n$ the vector of all ones, that is $e_i = 1$ for all $i = 1, 2, \ldots, n$.

We write $A \geq 0$ and say A is non-negative if all its elements $A_{i,j}$ are non-negative. We say that a symmetric matrix is *reducible* if there exists a permutation $\pi \in \mathcal{P}_n$ such that

$$A^\pi = \begin{pmatrix} B & 0 \\ 0 & C \end{pmatrix}, \tag{3.1.1}$$

where B and C are non empty square matrices. If no such permutation exists then A is *irreducible*. If there exists a permutation π and square matrices $A_1, A_2, \ldots, A_k$ which are irreducible for some k, such that:

$$A^\pi = \begin{pmatrix} A_1 & 0 & \ldots & 0 \\ 0 & A_2 & (0) & \vdots \\ \vdots & (0) & \ddots & \vdots \\ 0 & \ldots & \ldots & A_k \end{pmatrix} \tag{3.1.2}$$

then we refer to $A_1, A_2, \ldots, A_k$ as the irreducible blocks of A.

We define the *Laplacian* of a symmetric matrix A to be $L_A = D_A - A$, where D_A is the degree matrix, which is a diagonal matrix with $d_{ii} = \sum_{j=1}^{n} a_{ij}$.

Definition 3.1.1 (Fiedler value, vector)**.** *Let A be a symmetric matrix, the Fiedler value is the smallest non-zero eigenvalue of its Laplacian matrix $L_A = D_A - A$, and the corresponding eigenvector is called Fiedler vector, using Rayleigh quotient the Fiedler vector can be seen as the optimal solution to*

$$\min\{y^T L_A y : y \in \mathbb{R}^n, y^T e = 0, \|y\|_2 = 1\}. \tag{3.1.3}$$

We recall the definition of $\mathcal{R}$-matrices presented in the introduction chapter,

Definition 3.1.2 ($\mathcal{R}$-matrices)**.** *We say that a matrix $A \in \mathbb{R}^{n\times n}$ is a $\mathcal{R}$-matrix (or Robinson matrix), if and only if, it is symmetric and satisfies the following property:*

$$i < j < k \Rightarrow A_{ij} \geq A_{ik} \text{ and } A_{jk} \geq A_{ik} \tag{3.1.4}$$

for all $1 \leq i, j, k \leq n$.

This definition is equivalent to say that, the inputs of the matrix A are increasing as we move toward the main diagonal,

Definition 3.1.3 (pre-$\mathcal{R}$ matrices)**.** *We say that a matrix $A \in \mathbb{R}^{n\times n}$ is a pre-$\mathcal{R}$ if there exists a permutation $\pi \in \mathcal{P}_n$ such that the matrix A^π is a $\mathcal{R}$-matrix. This is equivalent to say that there exists a permutation matrix Π for which $\Pi A \Pi^T$ is a $\mathcal{R}$-matrix. Figure 1.2 shows the structure of $\mathcal{R}$-matrices.*

Example 3.1.1. *Define the matrices A, B and Π to be,*

$$A = \begin{pmatrix} 5 & 1 & 0 & 2 & 4 & 4 \\ 1 & 2 & 1 & 2 & 1 & 0 \\ 0 & 1 & 1 & 0 & 0 & 0 \\ 2 & 2 & 1 & 3 & 2 & 1 \\ 4 & 1 & 0 & 2 & 4 & 3 \\ 4 & 0 & 0 & 1 & 3 & 6 \end{pmatrix}, \; B = \begin{pmatrix} 1 & 1 & 0 & 0 & 0 & 0 \\ 1 & 2 & 2 & 1 & 1 & 0 \\ 1 & 2 & 3 & 2 & 2 & 1 \\ 0 & 1 & 2 & 4 & 4 & 3 \\ 0 & 1 & 2 & 4 & 5 & 4 \\ 0 & 0 & 1 & 3 & 4 & 6 \end{pmatrix}, \; \Pi = \begin{pmatrix} 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

then A is clearly a $\mathcal{R}$-matrix because the entries are increasing as we move toward the diagonal, and B is a pre-$\mathcal{R}$ matrix. In fact, Π is a permutation matrix that corresponds to the permutation $\pi = \left(\begin{smallmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ 5 & 2 & 1 & 3 & 4 & 6 \end{smallmatrix}\right)$, and $B^\pi = \Pi B \Pi^T = A$.

Given n elements $x_1, x_2, \ldots, x_n$ in a data space $\mathcal{X}$, and a similarity function $f(\cdot,\cdot) : \mathcal{X}^2 \to \mathbb{R}^+$ that measures the desire for each pair of element x_i and x_j to be near to each other, that is $f(x_i, x_j)$ is large when x_i and x_j are close to each other and $f(x_i, x_j)$ is small when x_i and x_j are far from each other. An example of a similarity function is $f(x, y) = \exp(-\|x - y\|^2 / 2\sigma^2)$ for some $\sigma > 0$, which is also known as the Gaussian kernel.

Assume that $x_1, x_2, \ldots, x_n$ can be ordered along a chain, where the similarity between variables decreases with their distance within this chain. Then *the seriation problem's* goal is to reconstruct this linear ordering given the pairwise similarities. Mathematically, the seriation problem seeks to find all the permutations π of the set $\{1, 2, \ldots, n\}$ such that if $i < j < k$ then $f(x_{\pi(i)}, x_{\pi(j)}) \geq f(x_{\pi(i)}, x_{\pi(k)})$ and $f(x_{\pi(j)}, x_{\pi(k)}) \geq f(x_{\pi(i)}, x_{\pi(k)})$, see Figure 1.1. In other words, we are trying to find the permutations π for which the elements $x_{\pi(1)}, x_{\pi(2)}, \ldots, x_{\pi(n)}$ form a linear order. We consider the matrix $A \in \mathbb{R}^{n \times n}$ defined by $A_{ij} = f(x_i, x_j)$ for all $i, j = 1, 2, \ldots, n$, we call A a *similarity matrix* (sometimes also called *correlation matrix*).

We can formulate the seriation problem in terms of $\mathcal{R}$-matrices, as follows:

• **The seriation problem:** Given a pre-$\mathcal{R}$ matrix A find all the permutations π for which the matrix A^π is a $\mathcal{R}$-matrix.

In what follows we will show how to solve this problem efficiently using the spectral seriation algorithm introduced by Atkins et al. in [1]. As mentioned in [1] we can see this problem as a discrete optimization problem, where the goal is to find all permutations π of the rows and columns of A which minimizes a given loss objective function or maximizes a merit objective function. A classical seriation criterion is the 2-SUM problem (loss function), which consists of the following program:

$$\begin{aligned} &\text{Minimize } \sum_{i=1}^{n}\sum_{j=1}^{n} A_{ij}(\pi_i - \pi_j)^2 \\ &\text{Subject to } \pi \in \mathcal{P}_n \end{aligned} \tag{3.1.5}$$

Other seriation criterions can be found in [27]. Unfortunately solving (3.1.5) is NP-hard due to the discrete nature of permutations. Instead we consider the following problem which is the continuous version of the 2-SUM problem (3.1.5) and whose solution can be used as a heuristic,

$$\begin{aligned} &\text{Minimize } \sum_{i=1}^{n}\sum_{j=1}^{n} A_{ij}(x_i - x_j)^2 \\ &\text{Subject to } x \in \mathbb{R}^n \text{ and } \sum_{i=1}^{n} x_i = 0 \text{ and } \sum_{i=1}^{n} x_i^2 = 1. \end{aligned} \tag{3.1.6}$$

The same idea was used for the spectral clustering algorithm (see e.g. [8]). If x solves problem (3.1.6), then we can sort it and thus find a permutation π such that $x_{\pi(1)} \leq x_{\pi(2)} \leq \cdots \leq x_{\pi(n)}$ then heuristically π solves problem (3.1.5). Note that

$$x^T L_A x = \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} A_{ij}(x_i - x_j)^2 \tag{3.1.7}$$

therefore a solution to problem (3.1.6) is just the Fiedler vector defined previously (see equation (3.1.3)). This heuristic approach suggests the following algorithm, which is the main algorithm of this thesis:

Algorithm 1: The spectral seriation algorithm.

Input : A pre-$\mathcal{R}$ matrix $A \in \mathbb{R}^{n\times n}$.

1 Compute the Fiedler vector x of the Laplacian matrix L_A.

2 Sort the Fiedler vector by finding π such that x^π is monotone.

Output: A permutation π such that A^π is $\mathcal{R}$-matrix.

In what follows we will show more rigorously that this heuristic approach holds true. Additionally, we will deal with the case of degenerate Fiedler vector which produces many permutations.

Before diving in the main results presented in [1], let's introduce the Perron-frobenius theorem which will be needed later on

Theorem 3.1.2 (Perron–Frobenius)**.** *Let $M \in \mathbb{R}^{n\times n}$ such that $M_{ij} \geq 0$ for all i, j. Define $\rho(M) = \max_i |\lambda_i(M)|$ to be the spectral radius of M, then*

1. *$\rho(M)$ is an eigenvalue of M, and*
2. *there is a non-zero vector $x \geq 0$ such that $Mx = \rho(M)x$*

We are now ready to present the main theorem of the paper [1]

Theorem 3.1.3. *If A is a $\mathcal{R}$-matrix then it has a monotone Fiedler vector.*

Sketch of the proof. The idea of the proof is the following, consider a vector $x = (x_1, \ldots, x_n)^T$ then put $x' = (x_2 - x_1, \ldots, x_n - x_{n-1})^T$ the vector of differences between neighboring inputs in x. Note that x is non-decreasing if and only if $x' \geq 0$, therefore it is enough to construct a matrix $M \geq 0$ and choose x' as in the Perron-Frobenius theorem. The authors in [1] introduced two matrices $S \in \mathbb{R}^{(n-1)\times n}$ and $T \in \mathbb{R}^{n\times(n-1)}$ that satisfy $Sx = x'$ and $ST = I_n$ and $TS = I_n - ee_1^T$. Then define $M_A = SL_AT = \{m_{i,j}\}$, where L_A is the Laplacian of A. They proved the following equivalence

$$L_Ax = \lambda x, x \neq \alpha e \Leftrightarrow M_Ax' = \lambda x', \text{ where } x' = Sx \neq 0.$$

Which shows that, the eigenvalues of M_A are the same as the eigenvalues of L_A with the zero eigenvalue removed, and the eigenvectors of M_A are differences between neighboring entries of the corresponding eigenvectors of L_A. Additionally, it was shown that $m_{i,j} < 0$ for all $i \neq j$ using the fact that A is a $\mathcal{R}$-matrix. Thus if we choose β to be a value greater than $\max_i\{\lambda_i, m_{i,i}\}$, where λ_i are the eigenvalues of M_A. Then $\tilde{M}_A = \beta I - M_A$ is nonnegative with eigenvalues $\tilde{\lambda} = \beta - \lambda_i$. Also, $\tilde{M}_A$ and M_A share the same set of eigenvectors. By Theorem 3.1.2, there exists a nonnegative eigenvector x' of $\tilde{M}_A$ corresponding to the largest eigenvalue of $\tilde{M}_A$. But x' is also an eigenvector of M_A corresponding to M_A's smallest eigenvalue. And this is just Sx, where x is a Fiedler vector of L_A. Since $x' = Sx$ is nonnegative, the corresponding Fiedler vector of L_A is nondecreasing and the theorem follows. (Note that since the sign of an eigenvector is unspecified, the Fiedler vector could also be increasing). ∎

Theorem 3.1.4. *Let A be a pre-$\mathcal{R}$ matrix with a simple Fiedler value and a Fiedler vector with no repeated values. Let π_1 (resp. π_2) be the permutation induced by sorting the values in the Fiedler vector in increasing (decreasing) order. Then A^{π_1} and A^{π_2} are $\mathcal{R}$-matrices, and no other permutations of A produce $\mathcal{R}$-matrices.*

Theorem 3.1.4 provides the essence of the Algorithm 1, in fact if A is a pre-$\mathcal{R}$ matrix with a simple Fiedler value and a Fiedler vector with no repeated values then the Algorithm 1 returns exactly the two permutations as in Theorem 3.1.4. Which explains the heuristic approach explained previously.

Now, what if the Fiedler vector is degenerate? i.e., it contains repeated values, in this case there is more than one permutation that makes the Fiedler vector non-decreasing for example. Are all those permutations going to transform A to a $\mathcal{R}$-matrix? The answer is yes, the authors of [1] proved that even in the case of repeated values the algorithm 1 holds true. Moreover, they proposed a detailed algorithm to find all the permutation π such that A^{π} in terms of PQ-trees. Theorem 3.1.4 requires strong assumptions about the spectrum of A, In what follows we will present a way to remove this restrictions.

Previously, we presented an algorithm to solve the seriation problem, unfortunately Theorem 3.1.4 lacks generality and require strong assumptions about the spectrum of the matrix A. We will generalize the Theorem 3.1.4 referring to results from [1]. Note that the proofs in [1] are very technical thus for the sake of our thesis we will note prove most of the theorems. In order to represent all the permutations that solves the seriation problem we will use a data structure called PQ-tree, thus we will start by a brief introduction to this data structure.

A PQ-tree is a tree-based data structure that represents a family of permutations on a set of elements. More formally a PQ-tree over a set $U = \{u_1, u_2, \ldots, u_n\}$ is a rooted, ordered tree in which each element is represented as an internal node (or leaf) and each non-leaf node is labelled P or Q. A P node has at least two children, and a Q node has at least three children.

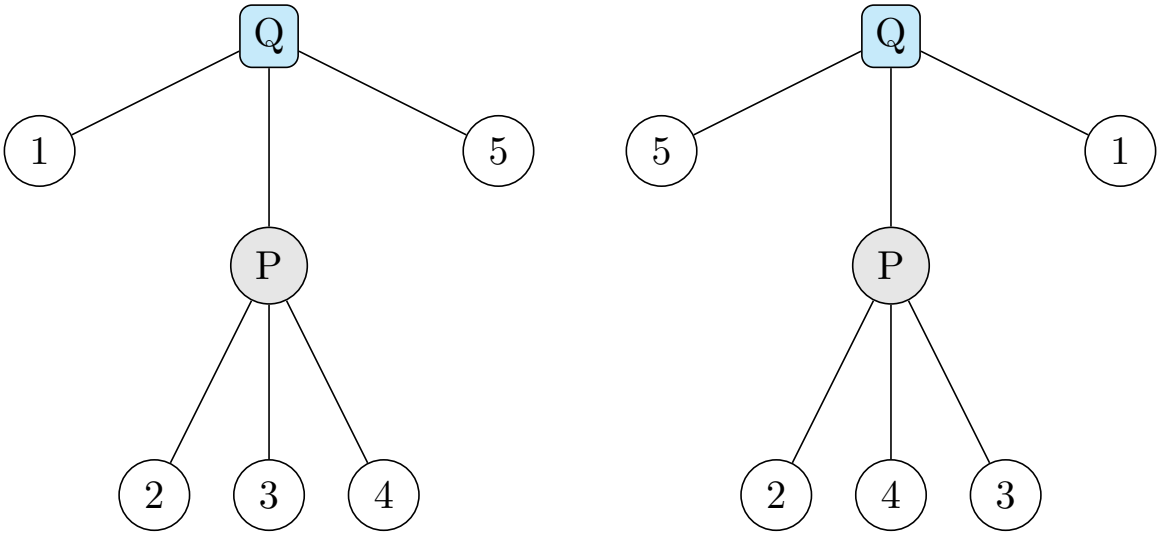
A PQ-tree represents its permutations via permissible re-orderings of the children of its nodes. In fact we say that two PQ-trees are equivalent if one can be transformed into the other by applying a sequence of the following two equivalence transformations.

1. Arbitrarily permute the children of a P-node.

2. Reverse the children of a Q-node.

Conveniently, the equivalence class represented by a PQ-tree corresponds precisely to the set of permutations consistent with an instance of a seriation problem. In what follows we will describe the general spectral seriation algorithm by [1] which uses Laplacian eigenvectors to construct a PQ-tree for an instance of the seriation problem.

Figure 3.1 shows an example of two equivalent PQ-trees. This PQ-tree represents the following twelve permutations on the set $\{1, 2, 3, 4, 5\}$: 12345, 12435, 13245, 13425, 14235, 14325, 52341, 52431, 53241, 53421, 54231, 54321.



(a) PQ-tree [1 (2 3 4) 5].

(b) PQ-tree [5 (2 4 3) 1].

Figure 3.1: Equivalent PQ-trees of the set $\{1, 2, 3, 4, 5\}$.

Now let's continue our discussion about the seriation problem. Several observations about the seriation problem will simplify the analysis. Note first that adding a coefficient to the similarity matrix A will not change the Fiedler vector. Thus, we might assume that the smallest value of the similarity matrix is zero. Moreover, note that substracting the smallest value from all the inputs of the matrix does not change the pre-$\mathcal{R}$ nature of the matrix. This can be stated as follows:

Lemma 3.1.5. *Let A be a symmetric matrix and let $\bar{A} = A - \alpha ee^T$ for some real α. A vector x is a Fiedler vector of A if and only if x is a Fiedler vector of $\bar{A}$. So without loss of generality we can assume that the smallest off-diagonal entry of A is zero.*

Proof. By the definition of a Laplacian it follows that $L_{\bar{A}} = L_A + \alpha ee^T - \alpha nI$, where n is the dimension of A. Then $L_{\bar{A}}e = 0$, but for any other eigenvector of x of L_A, $L_{\bar{A}}x = L_A x + 0 - \alpha nx$. That is, the eigenvalues are simply shiften down by αn while the eigenvectors are preserved. ∎

This will justify the first step of the general algorithm 2, presented by [1], which makes the smallest off-diagonal value of our pre-$\mathcal{R}$ matrix A to be 0. Next we can decouple the seriation problem to the irreducible blocks of the matrix A. In fact, we can solve the seriation problem for each irreducible block of A and concatenate the permutation solutions to get a permutation that makes A a $\mathcal{R}$-matrix. More formally, we have the following lemma.

Lemma 3.1.6. *Let A_i, $i = 1, \ldots, k$, be the irreducible blocks of a pre-$\mathcal{R}$-matrix A, and let π_i be a permutation of block A_i such that the submatrix $A_i^{\pi_i}$ is a $\mathcal{R}$-matrix. Then any*

permutation formed by concatenating the π_i'x will make A become a $\mathcal{R}$-matrix. In terms of PQ-tree, the π_i permutations are children of a single P-node.

With these preliminary results, from here on we will assume that the smallest off-diagonal value is zero and that the matrix is irreducible.

Lemma 3.1.7. *Let A be a $n \times n$ $\mathcal{R}$-matrix with a monotone Fiedler vector x. If $\mathcal{J} = [r, s]$ is a maximal interval such that $x_r = x_s$, then for any $k \notin \mathcal{J}$, $a_{r,k} = a_{r+1,k} = \cdots = a_{s,k}$.*

The proof of this Lemma can be found in [1], and its main idea is to subtract lines s and r on the equation $L_A x = \lambda x$ and use the fact that x is monotone as well as the $\mathcal{R}$-property of A. Now we will present a Lemma that is essentially the converse of the previous Lemma, but we will not discuss its proof here.

Lemma 3.1.8. *Let A be an irreducible $n \times n$ matrix with $a_{n,1} = 0$. If $\mathcal{J} = [r, s] \neq [1, n]$ is an interval such that $a_{r,k} = a_{s,k}$ for all $k \notin \mathcal{J}$, then $x_r = x_{r+1} \cdots x_s$ for any Fiedler vector x.*

Lemma 3.1.9. *Let A be an irreducible $\mathcal{R}$-matrix with $a_{n,1} = 0$, and x, a monotone Fiedler vector of A. If $\mathcal{J} = [r, s]$ is an interval such that $x_r = x_{r+1} = \cdots = x_s$, then for any Fiedler vector y, $y_r = y_{r+1} = \cdots = y_s$.*

Now we present and prove a very important Lemma about the Fiedler value:

Lemma 3.1.10. *If A is an irreducible $\mathcal{R}$-matrix with $a_{n,1} = 0$, then the Fiedler value λ_2 is a simple eigenvalue.*

Proof. Assume that λ_2 is not simple, then take two linearly independent Fiedler vectors x and y, with x non-decreasing. Define

$$k := \arg\max \left\{ \frac{y_{j+1} - y_j}{x_{j+1} - x_j} \text{ s.t. } j \in \{1, 2, \ldots, n\} \text{ and } x_{j+1} \neq x_j \right\},$$

such k should exists since $x^T e = 0$. Now define $z = (y_{k+1} - y_k)x - (x_{k+1} - x_k)y$, note that z is an other Fiedler vector because x and y are linearly independent. Let $j \in \{1, 2, \ldots, n\}$, if $x_{j+1} = x_j$ then by Lemma 3.1.8 $y_{j+1} = y_j$ thus $z_{j+1} = z_j$ by definition of z. Moreover if $x_{j+1} \neq x_j$ then

$$z_{j+1} - z_j = (x_{j+1} - x_j)(x_{k+1} - x_k) \left(\frac{y_{k+1} - y_k}{x_{k+1} - x_k} - \frac{y_{j+1} - y_j}{x_{j+1} - x_j} \right). \tag{3.1.8}$$

then $z_{j+1} - z_j \geq 0$ by definition of k. Moreover $z_{k+1} = z_k$, therefore by Lemma 3.1.8 the indices of any repeated values in z must be repeated in x that is $x_{k+1} = x_k$ which contradicts the definition of k. We conclude that λ_2 should be simple. ∎

Finally, we need to deal with the case where the Fiedler vector has repeated values. As the following theorem states, repeated values decouple the problem into pieces that can be solved in a recursive way.

Theorem 3.1.11. *Let A be a pre-$\mathcal{R}$-matrix with a simple Fiedler value and Fiedler vector x. Suppose there is some repeated value β in x and define*

1. $\mathcal{I} := \{1 \leq i \leq n \text{ s.t. } x_i < \beta\}$,
2. $\mathcal{J} := \{1 \leq i \leq n \text{ s.t. } x_i = \beta\}$ *and*
3. $\mathcal{K} := \{1 \leq i \leq n \text{ s.t. } x_i > \beta\}$.

Then π is a $\mathcal{R}$-matrix ordering for A if and only if π or its reversal can be expressed as (π_i, π_j, π_k), where π_j is a $\mathcal{R}$-matrix ordering for the submatrix $A(\mathcal{J}, \mathcal{J})$ of A induced by $\mathcal{J}$, and π_i and π_k are the restrictions of some $\mathcal{R}$-matrix ordering for A to $\mathcal{I}$ and $\mathcal{K}$, respectively.

With all these results together, we are now ready to present the main algorithm 2 of [1] which returns, a PQ-tree, that contain all the permutations π that solve the seriation problem for a given pre-$\mathcal{R}$ matrix A.

Algorithm 2: Spectral-Sort algorithm [1].

Input : A pre-$\mathcal{R}$ matrix $A \in \mathbb{R}^{n \times n}$.
U, a set of indices for the rows/columns of A.

Output: T, a PQ-tree that encodes the set of all permutations π such that A^π is a $\mathcal{R}$-matrix.

$\alpha := \min_{i \neq j} a_{i,j}$
$A := A - \alpha e e^T$
$\{A_1, A_2, \ldots, A_k\} :=$ the irreducible blocks of A
$\{U_1, U_2, \ldots, U_k\} :=$ the corresponding index sets
if $k > 1$ **then**
  for $j := 1, 2, \ldots, k$ **do**
    $T_j := \text{Spectral} - \text{Sort}(A_j, U_j)$
  end
  $T := \text{P} - \text{node}(T_1, T_2, \ldots, T_k)$
else
  if $n = 1$ **then**
    $T := u_1$
  else if $n = 2$ **then**
    $T := \text{P} - \text{node}(u_1, u_2)$
  else
    $x :=$ Fiedler vector for L_A
    Sort x
    $t :=$ number of distinct values in x
    for *j := 1,2,..., t* **do**
      $V_j :=$ indices of elements in x with j th value
      $T_j :=$ Spectral-Sort$(A(V_j, V_j), V_j)$
    end
    $T := \text{Q} - \text{node}(T_1, \ldots, T_T)$
  end
end

3.2 The consecutive ones problem

In chapter 2 we presented the archaeological problem of "sequence dating" formulated by Flinders Petrie in his paper [23], which corresponds to sequencing chronologically a set of graves given some artifacts, which may or may not be present in the graves. This can be formulated mathematically in terms of $(0,1)$-matrices, we are given a $n \times m$ $(0,1)$-matrix A, where the rows of the matrix corresponds to graves and columns represents the artifacts, and $a_{i,j} = 0$ or 1 according to whether the artifact j is found in the grave i or not. The problem of sequence dating is to permute the rows of the matrix so as to make the ones in each column appear consecutively. We shall call this problem *The consecutive ones problem*

or C1P for short.

In this section we will briefly describe a graph algorithm method to solve the C1P proposed by Fulkerson and Gross in [13] and discuss an application to the recognition of interval graphs. We will also present some theorems by Kendall in [17, 18] that shows the equivalence between the seriation problem and the C1P and the recognition of unimodal matrices.

3.2.1 A graph algorithm for the C1P

For the sake of the consecutive ones problem, we define a class of matrices:

Definition 3.2.1 ($\mathcal{P}$-matrices)**.** *We say that a $n \times m$ $(0,1)$-matrix A is a Petrie matrix or $\mathcal{P}$-matrix if the ones in each column of A appear consecutively. Moreover, we say that A is pre-$\mathcal{P}$ matrix if there exists a permutation matrix Π such that ΠA is a $\mathcal{P}$-matrix (we also say that A has the consecutive one property).*

Example 3.2.1. *Define the matrices A, B and Π to be,*

$$A = \begin{pmatrix} 1 & 0 & 0 & 0 & 1 \\ 1 & 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 1 & 0 \\ 0 & 1 & 1 & 0 & 0 \end{pmatrix}, \; B = \begin{pmatrix} 1 & 0 & 1 & 1 & 0 \\ 1 & 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 1 & 0 \\ 1 & 0 & 0 & 0 & 1 \\ 0 & 1 & 1 & 0 & 0 \end{pmatrix}, \; \Pi = \begin{pmatrix} 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

then A is clearly a $\mathcal{P}$-matrix because the ones in each column are consecutive, and B is a pre-$\mathcal{P}$ matrix. In fact, Π is a permutation matrix that corresponds to the permutation $\pi = \left(\begin{smallmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ 5 & 2 & 1 & 3 & 4 & 6 \end{smallmatrix}\right)$, and $\Pi B = A$.

In the previous example the permutation matrix Π was given, but if the number of rows m is large, then finding the latent permutation naively requires $m!$ steps, which is huge computationally! Given a $(0,1)$-matrix A of size $n \times m$, the authors of the paper [13] showed that we can answer the question of whether there is a rearrangement of the rows that will make A a $\mathcal{P}$-matrix if we merely knew $A^T A$. Moreover, they proposed a method that takes as input $A^T A$, and requires $\mathcal{O}(n^2)$ to solve this problem. Here we will discuss their method, and in the next subsection we show an application to interval graphs.

Given a $m \times n$ $(0,1)$-matrix A. An important question that the authors in [13] posed is: how much information about A is needed to decide whether A is a pre-$\mathcal{P}$ matrix or not? Do we need A itself or something less suffice? Theorem 3.2.2 provides a partial answer; it states that a knowledge of $A^T A$ is enough. The proof can be found in [13].

Theorem 3.2.2. *Let A and B be two $(0,1)$-matrices satisfying*

$$A^T A = B^T B. \tag{3.2.1}$$

Then either both A and B are pre-$\mathcal{P}$ matrices or neither is. Moreover if A and B have the same number of rows and A is a pre-$\mathcal{P}$ matrix then there is a permutation matrix Π such that $B = \Pi A$.

Now we present and prove an immediate corollary:

Corollary 3.2.3. *Let A and B be two $(0,1)$-matrices satisfying equation (3.2.1). If A is a pre-$\mathcal{P}$ matrix and has no rows of zeros, then there exists a permutation matrix Π such that*

$$\Pi B = \begin{pmatrix} A \\ 0 \end{pmatrix}$$

Proof. Assume that $A \in \{0,1\}^{m_1 \times n}$ and $B \in \{0,1\}^{m_2 \times n}$, if $m_2 < m_1$ then we can add $m_1 - m_2$ rows of 0's to B, and construct the matrix $C = \begin{pmatrix} B \\ 0 \end{pmatrix}$, clearly $C^T C = A^T A$. Moreover, A and C have the same number of rows, then by theorem 3.2.2 A is a row permutation of C which is a contradiction since A has no rows of 0's. Therefore, $m_2 \geq m_1$, the conclusion follows by adjoining $m_2 - m_1$ rows of 0's to A. ∎

Definition 3.2.2. *Let $a, b \in \{0,1\}^m$ then it is clear that*

$$0 \leq a \cdot b \leq \min(a \cdot a, b \cdot b).$$

We say that

- *a and b overlap if $0 < a \cdot b < \min(a \cdot a, b \cdot b)$.*
- *a and b are disjoint if $a \cdot b = 0$.*
- *a contains b if $a \cdot b = b \cdot b$.*

Definition 3.2.3. *We say that a $m \times n$ $(0,1)$-matrix A with columns $a_1, a_2, \ldots a_n$ is proper if $a_i \neq 0$ for all $i = 1, 2, \ldots, m$ and $a_i \neq a_j$ for all $i \neq j$.*

When studying the consecutive one property of a matrix it is convenient to impose without loss of generality that the matrix is proper. We will now define some graphs associated with $(0,1)$-matrices,

Definition 3.2.4. *Given a $m \times n$ $(0,1)$-matrix A with columns $a_1, a_2, \ldots, a_n$ define:*

- *The full graph $\mathscr{F}(A)$: consider vertices $v_1, v_2, \ldots, v_n$ corresponding to the columns of A. Put an undirected edge $\{v_i, v_j\}$ joining v_i and v_j if the columns a_i and a_j overlaps; and a directed edge $[v_i, v_j]$ from v_i to v_j if the column a_i contains a_j.*
- *The overlap graph $\mathscr{C}(A)$: it is a subgraph of $\mathscr{F}(A)$ where the edges corresponds only to overlapping columns.*
- *The component graph $\mathscr{D}(A)$: it is obtained from overlap graph $\mathscr{C}(A)$ as follows; we consider its vertices $X_1, X_2, \ldots, X_p$ to be the connected components of $\mathscr{C}(A)$. We put a directed edge $[X_i, X_j]$ from X_i to X_j if there exists some $v_i \in X_i$ and $v_j \in X_j$ such that $[v_i, v_j]$ is an edge from v_i to v_j in $\mathscr{F}(A)$.*

Note that for a matrix A we can construct the graphs $\mathscr{F}(A)$, $\mathscr{C}(A)$ and $\mathscr{D}(A)$ if we know A^TA, since all these graphs are based on the pairwise dot product of the columns of A. Moreover, for each two columns a_i and a_j in A the value of the dot product $a_i \cdot a_j$ is given in the position (i, j) in the matrix A^TA. To illustrate the concept of these graphs let's give an example from [13]:

Example 3.2.4. *Let*

$$A = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 & 1 & 0 & 1 \\ 1 & 1 & 1 & 0 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 & 1 & 1 & 0 \\ 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 & 0 & 0 & 1 \end{pmatrix} \tag{3.2.2}$$

then

$$A^TA = \begin{pmatrix} 6 & 5 & 5 & 0 & 0 & 2 & 2 & 3 \\ 5 & 8 & 5 & 2 & 2 & 2 & 2 & 3 \\ 5 & 5 & 5 & 0 & 0 & 2 & 2 & 3 \\ 0 & 2 & 0 & 2 & 1 & 0 & 0 & 0 \\ 0 & 2 & 0 & 1 & 2 & 0 & 0 & 0 \\ 2 & 2 & 2 & 0 & 0 & 2 & 1 & 1 \\ 2 & 2 & 2 & 0 & 0 & 1 & 2 & 0 \\ 3 & 3 & 3 & 0 & 0 & 1 & 0 & 3 \end{pmatrix}.$$

To construct the full graph, we consider eight vertices $v_1, v_2, \ldots, v_8$, then put directed edges from v_i to v_j if the column a_i contains the column a_j, i.e., $a_i \cdot a_j = a_j \cdot a_j$. For example $a_1 \cdot a_3 = a_3 \cdot a_3 = 5$, thus we put a directed edge from a_1 to a_3. We also put undirected edges between two vertices if they overlap, that is if two columns a_i and a_j satisfy $0 < a_i \cdot a_j < \min(a_i \cdot a_i, a_j \cdot a_j)$. for example columns a_1 and a_2 overlap because $0 < a_1 \cdot a_2 = 5 < \min(a_1 \cdot a_1, a_2 \cdot a_2) = \min(6, 8) = 6$. The overlap graph $\mathscr{C}(A)$ is then given just by removing the directed edges. The two graphs are given in Figure 3.3 and Figure 3.2.

It is clear that the connected components of the graph $\mathscr{C}(A)$ are $X_1 = \{v_1, v_2\}$, $X_2 = \{v_3\}$, $X_3 = \{v_4, v_5\}$ and $X_4 = \{v_6, v_7, v_8\}$. To construct the component graph $\mathscr{D}(A)$; we consider the vertices X_1, X_2, X_3, X_4 we then put directed edges between X_i and X_j if there exists v_i and v_j in X_i and X_j respectively such that there is a directed edge from v_i to v_j in the full graph $\mathscr{F}(A)$, the component graph is shown in Figure 3.4.

Now we will present some results from [13] about the graphs defined previously which will make the method for solving `C1P` clearer. We recall that a graph is said to be acyclic if it has no cycles. Moreover, we say that a graph is transitive if it has the following property; if $[u, v]$ is a directed edge from u to v and $[v, w]$ is an other directed edge from v to w in the graph, then $[u, w]$ is a directed edge from u to w.

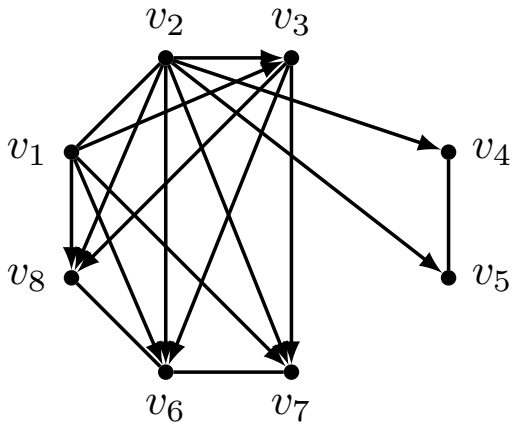


Figure 3.2: The full graph $\mathscr{F}(A)$ of the matrix in equation (3.2.2).

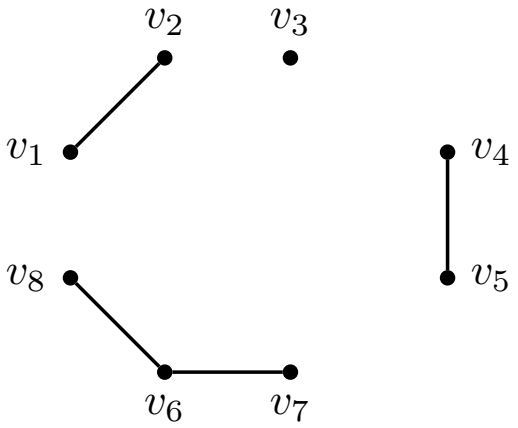


Figure 3.3: The overlap graph $\mathscr{C}(A)$ of the matrix in equation (3.2.2).

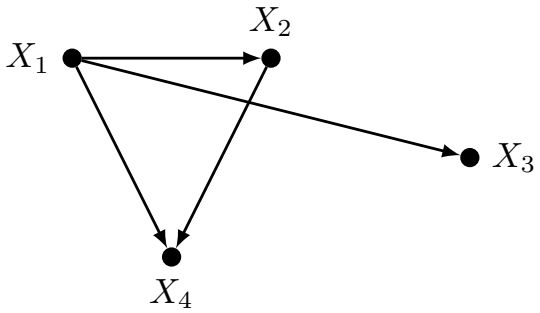


Figure 3.4: The component graph $\mathscr{D}(A)$ of the matrix in equation (3.2.2).

Theorem 3.2.5. *The component graph $\mathscr{D}(A)$ of a proper $(0,1)$-matrix A is acyclic and transitive.*

Figure 3.4 shows the component graph which is acyclic and transitive. Theorem 3.2.5 shows that the graph $\mathscr{D}(A)$ defines a partial order on its vertices. In fact, for each two vertices X and Y of $\mathscr{D}(A)$ define the order $\preceq$ as follows:

$$X \preceq Y \Leftrightarrow X = Y \text{ or } [Y, X] \text{ is an edge in } \mathscr{D}(A). \tag{3.2.3}$$

The fact that $\mathscr{D}(A)$ is acyclic and transitive ensures that $\preceq$ is a partial order (i.e., reflexive - antisymmetric - transitive). Moreover, this ordering is important in the sense that any vertex X can have at most one predecessor, as the following theorem shows:

Theorem 3.2.6. *Let A be a proper $(0,1)$-matrix and let X, Y, Z be three vertices of the connected component graph $\mathscr{D}(A)$. If $[X,Z]$ and $[Y,Z]$ are edges of $\mathscr{D}$, then either $[X,Y]$ or $[Y,X]$ is an edge. In other terms, each vertex of $\mathscr{D}$ has at most one predecessor.*

An other useful fact about the full graph $\mathscr{F}(A)$ of a proper $(0,1)$-matrix A is the following:

Theorem 3.2.7. *Let A be a proper $(0,1)$-matrix and let X and Y be two components of the overlap graph $\mathscr{C}(A)$. If there exists $x \in X$ and $y \in Y$ such that $[x,y]$ is an edge of $\mathscr{F}(A)$ then $[x,y']$ is an edge of $\mathscr{F}(A)$ for all $y' \in Y$.*

This theorem shows that each column of A corresponding to the component Y will be contained in some columns corresponding to the component X and disjoint from the others.

Given a $(0,1)$-matrix A, let $X_1, X_2, \ldots, X_p$ be the connected components of the overlap graph $\mathscr{C}(A)$. We can write in a unique way, $A = (A_1, A_2, \ldots, A_p)$ where each A_k is a submatrix of A corresponding to X_k. We refer to the A_k's the *overlap components* of A. Moreover, we say that A is connected if it has only one component.

Theorem 3.2.8. *Let A be a $(0,1)$-matrix, write $A = (A_1, A_2, \ldots, A_p)$ where the A_k's are the overlap components of A. Then A is a pre-$\mathcal{P}$ matrix if and only if A_k is a pre-$\mathcal{P}$ matrix for all $k = 1, 2, \ldots, p$.*

The above theorem shows that solving the `C1P` for a $(0,1)$-matrix can be decomposed into two stages. First, propose an algorithm that solve the `C1P` for a connected matrix. Second, propose a method to merge the solutions of the connected components of the matrix A.

The idea of solving the `C1P` for connected matrices will be based on *rigid triples* which are:

Definition 3.2.5. *Let $a, b, c \in \{0,1\}^m$ we say that (a,b,c) form a rigid triple if a overlaps b and b overlaps c.*

Let A be a pre-$\mathcal{P}$ connected matrix and let (a,b,c) be three columns of A that form a rigid triple. Then there exists a permutation matrix Π such that the matrix $B = \Pi A$ contains one of the following configurations:

- The configuration C_1: when $a \cdot c < \min(a \cdot b, b \cdot c)$, we start by writing down $a \cdot a$ consecutively in a column, then $b \cdot b$ consecutively in an other column so that the two columns share $a \cdot b$ number of ones. Then we place the third column c in the same way. For example if $a \cdot a = 5$, $b \cdot b = 6$, $c \cdot c = 4$, $a \cdot b = 4$, $b \cdot c = 3$ and $a \cdot c = 1$ then we have

the following configuration

a	b	c
1		
1	1	
1	1	
1	1	
1	1	1
	1	1
	1	1
		1

- The configuration C_2: when $a \cdot c \geq \min(a \cdot b, b \cdot c)$, this configuration is different from the last one only by placing the third column c on top of the second. For example, if $a \cdot a = 5$, $b \cdot b = 6$, $c \cdot c = 4$, $a \cdot b = 4$, $b \cdot c = 3$ and $a \cdot c = 4$ then we have the following configuration

a	b	c
1		1
1	1	1
1	1	1
1	1	1
1	1	
	1	
	1	

The construction of the latent $\mathcal{P}$-matrix for a connected pre-$\mathcal{P}$ matrix A will be based on rigid triples. The idea, is to remove edges from $\mathscr{C}(A)$ so as to make a spanning subtree, then positioning strings of ones using rigid triples obtained from this tree. Start with the root of the tree as the first component of a rigid triple, and position the triple respecting the configuration C_1 or C_2. Then, referring to $\mathscr{C}$ we add more columns relative to other rigid triples, at the end of this process we will form a $\mathcal{P}$-matrix B which is a row permutation of A by Theorem 3.2.2. Note that if the matrix has n columns this method requires $\mathcal{O}(n^2)$ steps. For more details see section 5 in [13].

Having proposed a method for solving the `C1P` for connected matrices, we need to propose a method that fit the components together. Let A be a pre-$\mathcal{P}$, proper $(0,1)$-matrix, and let $A_1, A_2, \ldots, A_p$ be the overlap components of A, consider the $\mathcal{P}$-matrix $B_1, B_2, \ldots, B_p$ to be the matrices constructed following the algorithm described previously for connected matrices. The idea of merging the solutions together is to use the partial ordering $\preceq$ over the vertices of $\mathscr{D}(A)$ (defined in equation (3.2.3)), then put the matrices B_k's in a decreasing order starting with the maximal elements of $\mathscr{D}$. Next delete the maximal elements from $\mathscr{D}(A)$ and look for the new maximal elements in the new reduced graph. In positioning the new maximal elements relative to the previous ones we refer to theorem 3.2.7. For more details see section 6 in [13].

3.2.2 Application to interval graphs

According to [13], the study of interval graphs was stimulated in part by a biological application concerning the fine structure of genes. A basic problem, is to decide whether or not the subelements of genes are linked together in a linear order. By conducting some experiments, we can gather data about the intersection of mutant genes and represent it as an undirected graph. If this graph is an interval graph, then the observed data is consistent with a linear order.

Definition 3.2.6 (Intersection graphs)**.** *Let $S_1, S_2, \ldots, S_n$ be a family of sets. The intersection graph $G = (V, E)$ corresponding to this family of sets is an undirected graph formed by creating one vertex v_i for each set S_i, i.e. $V = \{v_1, v_2, \ldots, v_n\}$ and connecting two vertices v_i and v_j by an edge whenever the corresponding two sets have a nonempty intersection, that is,*

$$E = \{\{v_i, v_j\} : S_i \cap S_j \neq \emptyset\}.$$

Conversely, any finite graph can of course be viewed as the intersection graph of a family of sets (in many ways).

Definition 3.2.7 (Interval graphs)**.** *If the family of sets $S_1, S_2, \ldots, S_n$ in the previous definition is an interval of $\mathbb{R}$, the associated intersection graph is called "interval graph". Figure 3.5 shows an example of an interval graph corresponding to seven real sets.*

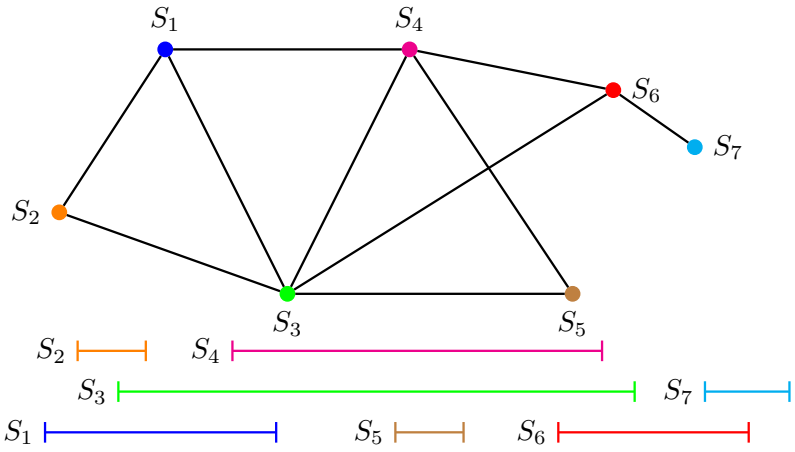


Figure 3.5: Seven intervals on the real line and the corresponding seven-vertex interval graph.

Note that unlike intersection graphs, there are some graphs which are not interval graphs. It was showed in [13] that recognizing interval graphs is related to the consecutive ones problem. This is done through a matrix that characterizes graphs, called dominant clique vs. vertex matrix. Given a graph G, consider the family of cliques of G and partially

order them by set inclusion, we call the maximal elements the dominant cliques of the graph. We then construct the dominant clique vs. vertex matrix D where the rows of D corresponds to the dominant cliques and columns to the vertices of G, and $d_{ij} = 1$ if the vertex j belongs to the dominant clique i, and $d_{ij} = 0$ otherwise. For the graph shown in Figure 3.6 the dominant cliques are $C_1 = \{a, c, e, g\}$, $C_2 = \{b, d, f, g\}$, $C_3 = \{f, g, h\}$ and $C_4 = \{e, g, h\}$. Then the dominant clique vs. vertex matrix is

$$D = \begin{array}{c} \begin{array}{cccccccc} a & b & c & d & e & f & g & h \end{array} \\ \begin{pmatrix} 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 1 & 0 & 1 & 1 \end{pmatrix} \end{array} \begin{array}{c} C_1 \\ C_2 \\ C_3 \\ C_4 \end{array} \tag{3.2.4}$$

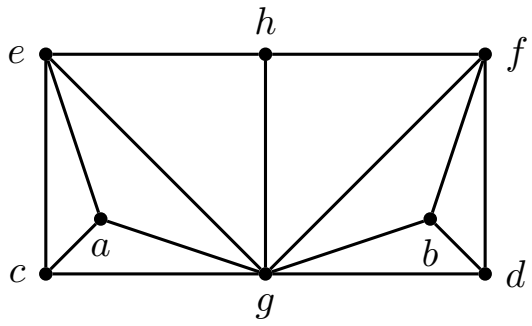


Figure 3.6

The theory and methods developed presented in [13] for the consecutive ones problem can be carried out to interval graphs via the following theorem

Theorem 3.2.9. *A graph G is an interval graph if and only if the dominant clique vs. vertex matrix D of G is a pre-$\mathcal{P}$ matrix.*

Clearly the matrix in equation (3.2.4) is a pre-$\mathcal{P}$ matrix (switch C_2 with C_4), thus the graph in Figure 3.6 is an interval graph.

3.2.3 The C1P and seriation problem

While the work of Fulkerson and Gross in [13] proved that the C1P for a $(0, 1)$-matrix A can be solved if we knew the product $A^T A$ (see Theorem 3.2.2), the work of Kendall in [17] show that the C1P can be solved if we know AA^T, that is both $A^T A$ and AA^T contain enough information to decide whether a matrix is a pre-$\mathcal{P}$ and to propose a sorting algorithm. Moreover, Kendall in [17] showed that there is a close relationship between the C1P and the seriation problem. The following results are due to D.G Kendall and proved in [17]

Lemma 3.2.10. *If A is a $\mathcal{P}$-matrix, then $S = AA^T$ is a $\mathcal{R}$-matrix.*

Lemma 3.2.11. *If A is a pre-$\mathcal{P}$ matrix and $S = AA^T$ is a $\mathcal{R}$-matrix then A is a $\mathcal{P}$-matrix.*

Theorem 3.2.12. *Let A be a pre-$\mathcal{P}$ matrix and put $S = AA^T$, let Π be a permutation matrix. Then ΠA is a $\mathcal{P}$-matrix if and only if $\Pi S \Pi^T$ is a $\mathcal{R}$-matrix.*

Let's give an example to illustrate this result, define the matrix B by

$$B = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 & 0 & 1 & 0 \\ 1 & 1 & 1 & 0 & 0 & 1 & 1 & 0 \\ 1 & 1 & 1 & 0 & 0 & 1 & 0 & 1 \\ 1 & 1 & 1 & 0 & 0 & 0 & 0 & 1 \\ 1 & 1 & 1 & 0 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 \end{pmatrix}, \tag{3.2.5}$$

then clearly B is a $\mathcal{P}$-matrix since all the ones appear consecutively. We have:

$$BB^T = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 \\ 1 & 4 & 4 & 3 & 3 & 3 & 1 & 1 & 1 \\ 1 & 4 & 5 & 4 & 3 & 3 & 1 & 1 & 1 \\ 1 & 3 & 4 & 5 & 4 & 4 & 1 & 1 & 1 \\ 1 & 3 & 3 & 4 & 4 & 4 & 1 & 1 & 1 \\ 1 & 3 & 3 & 4 & 4 & 4 & 1 & 1 & 1 \\ 0 & 1 & 1 & 1 & 1 & 1 & 2 & 2 & 1 \\ 0 & 1 & 1 & 1 & 1 & 1 & 2 & 3 & 2 \\ 0 & 1 & 1 & 1 & 1 & 1 & 1 & 2 & 2 \end{pmatrix}$$

which is a $\mathcal{R}$-matrix since the inputs of BB^T are increasing when moving toward the diagonal.

The method in [13] tells us whether a given $(0,1)$ matrix A has the consecutive ones property and finds a $\mathcal{P}$-matrix B that is a row permutation of A. Theorem 3.2.12 enables us to find all the permutations Π that will make ΠA a $\mathcal{P}$-matrix. In fact for a given pre-$\mathcal{P}$ matrix A, we can calculate $S = AA^T$ and use algorithm 2 for the matrix S to return all the permutations Π that will make S a $\mathcal{R}$-matrix. Theorem 3.2.12 ensures that those permutations are also solutions for the `C1P` problem for the matrix A. We conclude that the matrix AA^T contains enough information to actually find all the solutions to the consecutive ones problem, which is much stronger than the matrix A^TA used in [13].

An other related problem is that of unimodal matrices. We say that a vector $x \in \mathbb{R}^n$ is unimodal if x increases to a maximum and decreased, that is there is some $1 \leq p \leq n$ such that $x_1 \leq x_2 \leq \cdots \leq x_p$ and $x_p \geq x_{p+1} \geq \cdots \geq x_n$. We say that a matrix is a $\mathcal{Q}$-matrix if all its columns are unimodal, similarly we say that a matrix A is pre-$\mathcal{Q}$ matrix if there exists a permutation matrix Π such that ΠA is a $\mathcal{Q}$-matrix. Note that $\mathcal{P}$-matrices are a special case

of of $\mathcal{Q}$-matrices. Kendall in an other paper [18] generalized the work presented in [17], he proved that the seriation problem is also related to that of permuting the row of a matrix so as to make the columns unimodal. Specifically he proved that theorem 3.2.12 holds true for $\mathcal{Q}$-matrices if we replace the usual matrix product by the circular product defined by; Given two matrices $A \in \mathbb{R}^{n\times p}$ and $B \in \mathbb{R}^{p\times m}$ the circular product of A and B is the $A \circ B \in \mathbb{R}^{n\times m}$ defined by

$$(A \circ B)_{ij} = \sum_{k=1}^{p} \min(a_{ik}, b_{k,j})$$

The main theorem of [18] is the following

Theorem 3.2.13. *Let A be a pre-$\mathcal{Q}$ matrix and put $S = A \circ A^T$, the circular product of A. Let Π be a permutation matrix. Then ΠA is a $\mathcal{Q}$-matrix if and only if $\Pi S \Pi^T$ is a $\mathcal{R}$-matrix.*

Note that the circular product is exactly the usual matrix product for binary matrices, in fact if $a, b \in \{0, 1\}$ then $\min(a, b) = ab$. Therefore theorem 3.2.12 is just a special case of theorem 3.2.13. Figure 3.7 shows an example of a random $\mathcal{P}$-matrix A that was simulated using the following process; Choose $N = 100$ the number of rows and columns, then for each column choose two random numbers a and b from $\{1, 2, \ldots, n\}$ so that $a \leq b$ then put a string of ones from a to b. As Figure 3.8 shows, the product AA^T is clearly a $\mathcal{R}$-matrix since it is increasing toward the diagonal.

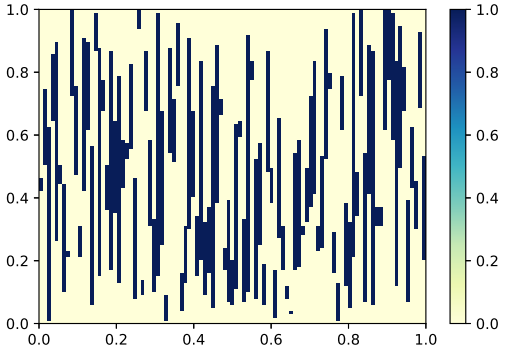


Figure 3.7: Example of a random $\mathcal{P}$-matrix A.

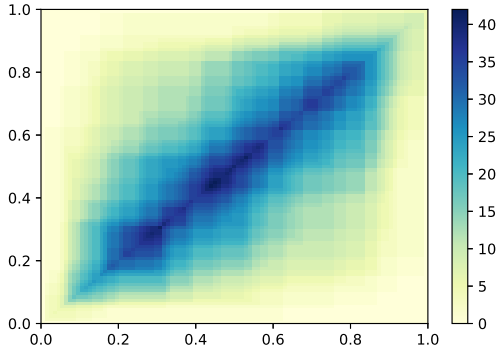


Figure 3.8: The product AA^T of the matrix in Figure 3.7.

Similarly Figure 3.9 shows an example of a random $\mathcal{Q}$-matrix that was simulated as follows; Construct a square matrix $N \times N$ with $N = 100$. Then for each column j, choose a random integer $p \in \{1, 2, \ldots, n\}$ then generate $u_1, u_2, \ldots, u_p$ uniformly from $[0, 1]$ and let $u_{(1)} \leq u_{(2)} \leq \cdots \leq u_{(p)}$ be the order statistics of the u_i's, then generate $u_{p+1}, \ldots, u_n$ uniformly from $[0, u_{(p)}]$ and similarly $u_{(p+1)} \geq u_{(p+2)} \geq \cdots u_{(n)}$ the order statistics (in a decreasing order); then put in column j the vector $[u_{(1)}, \ldots, u_{(p)}, \ldots u_{(n)}]^T$. The matrix in Figure 3.10 shows the circular product $A \circ A^T$ which is again a $\mathcal{R}$-matrix as Theorem 3.2.13 states.

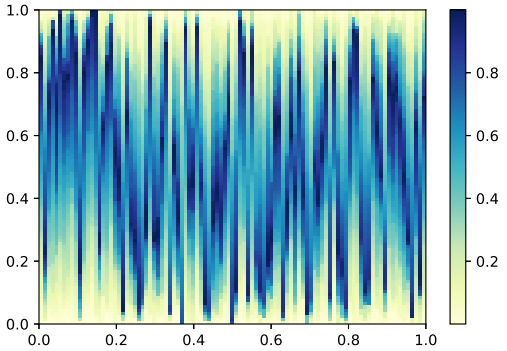


Figure 3.9: Example of a random $\mathcal{Q}$-matrix A.

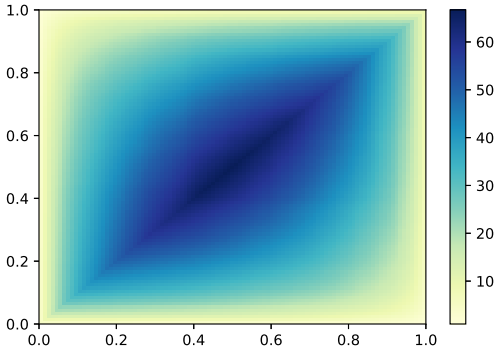


Figure 3.10: The circular $A \circ A^T$ of the matrix in Figure 3.9.

3.3 The quadratic assignment problem

The quadratic assignment problem is an other combinatorial problem, which is related to the seriation problem [20]. For a positive integer n, we are given n facilities, n locations, a flow matrix A whose entry $A_{i,j}$ represents the flow of activity between two facilities i and j, and a distance matrix B whose entry $B_{i,j}$ represents the distance between the locations i and j. Then the objective is to find an assignment of the facilities to the locations, i.e., a permutation $\pi \in \mathcal{P}_n$, minimizing the total cost of the assignment. The quadratic assignment problem `QAP`(A,B), for the matrices A and B, corresponds to solving the following optimization problem:

$$\begin{aligned} &\text{Minimize} \quad \sum_{1 \le i,j \le n} A_{i,j} B_{\pi(i),\pi(j)} \\ &\text{Subject to} \quad \pi \in \mathcal{P}_n. \end{aligned} \tag{3.3.1}$$

Note that the 2-SUM problem, that was introduced in equation (3.1.5) as a criterion for the seriation problem, is a special case of the quadratic assignment problem, in fact for the matrix B defined by $B_{i,j} = (i-j)^2$ for all $1 \le i, j \le n$, equation (3.3.1) becomes the same as equation (3.1.5).

A matrix A is said to be a Toeplitz matrix, if and only if, for all $1 \le i, j \le n-1$ we have $A_{i,j} = A_{i+1,j+1}$. The following theorems are proved in [20] which shows the relationship between the `QAP` and the seriation problem:

Theorem 3.3.1. *Let A and B be two $n \times n$ symmetric matrices and assume that A is a $\mathcal{R}$-matrix and that $-B$ is a $\mathcal{R}$-matrix and moreover A or B is a Toeplitz matrix. Then the identity permutation is an optimal solution to the problem `QAP`(A,B).*

Theorem 3.3.2. *Let A and B be two $n \times n$ symmetric matrices and assume that A is a pre-$\mathcal{R}$ matrix and that $-B$ is a pre-$\mathcal{R}$ matrix. Let π and τ be two permutations in $\mathcal{P}_n$ such*

that A^π and $-B^\tau$ are $\mathcal{R}$-matrices. Furthermore, assume that one of the matrices A^π or B^τ is a Toeplitz matrix. Then the permutation $\tau^{-1}\pi$ is optimal for the problem `QAP`*(A,B).*

It was also shown in [12] that the seriation problem is equivalent to the 2-SUM problem for a class of matrices. For a given symmetric matrix A, a generalization of the 2-SUM problem in equation (3.1.5) is

$$\begin{aligned} &\text{Minimize} \quad f(y_\pi) := \sum_{1\le i,j\le n} A_{i,j}(y_{\pi(i)} - y_{\pi(j)})^2 \\ &\text{Subject to} \quad \pi \in \mathcal{P}_n. \end{aligned} \tag{3.3.2}$$

where $y \in \mathbb{R}^n$ is a vector, note that for $y = (1, 2, \ldots, n)^T$ we get equation (3.1.5). This problem is related to the seriation problem via the following theorem from [12]

Theorem 3.3.3. *Let $A \in \mathbb{R}^{n\times m}$ be a pre-$\mathcal{Q}$ matrix and $y_i = ai + b$ for $i = 1, 2, \ldots, n$ for some $a, b \in \mathbb{R}$ where $a \neq 0$. Let $S = A \circ A^T$ be the circular product of A, if Π is a permutation matrix such that $\Pi S \Pi^T$ is a $\mathcal{R}$-matrix, then the corresponding permutation π solves the minimization combinatorial problem (3.3.2).*

Chapter 4

Noisy seriation

In chapter 2 we pointed out that, the seriation problem has many applications from its origins in archaeological studies to recent applications to ecology and sociology. In most of these applications, it is natural to expect the data to be noisy. In this case the reordered matrix may not be a perfect Robinsonian matrix, but it will be close to one. In the presence of noise, the goal of noisy seriation is to find an optimal permutation that makes the matrix "almost Robinsonian". In chapter 3 we presented two methods for solving the `C1P` and seriation problem, these algorithms applies only in the noise-free settings, that is when the input matrix is perfectly a pre-$\mathcal{R}$ matrix. Other efficient methods have been proposed in the literature see [21]. Surprisingly, despite decades of work on the seriation problem, little is known about the robustness to noise of such methods. In this chapter we will present some attempts for solving this problem.

4.1 An optimization parameter for noisy seriation

In this section we will summarize the work presented in the paper [16]. The authors introduced a new parameter $\Gamma_1(A)$ that measures how badly a matrix A fails to be a R-matrix, that is $\Gamma_1(A)$ is small when A is close to a $\mathcal{R}$-matrix, and precisely equal to 0 when it is a $\mathcal{R}$-matrix. This parameter is simply formulated and can be computed in linear time. It basically counts the number of local violations to the Robinson similarity property (see equation 3.1.4). Here we denote by $\mathcal{A}_n$ the set of $n \times n$ symmetric matrices with coefficients in the interval $[0, 1]$. Given $A \in \mathcal{A}_n$, the parameter $\Gamma_1(A)$ is defined as follows,

$$\Gamma_1(A) = \frac{1}{n^3} \sum_{1 \leq i < k < j \leq n} [A_{i,j} - A_{i,k}]_+ + [A_{i,j} - A_{k,j}]_+ \tag{4.1.1}$$

where $[x]_+ = x$ if $x \geq 0$, and 0 otherwise. It is straightforward to see that $\Gamma_1(A) = 0$ if and only if A is a R-matrix. Moreover $\Gamma_1(A) \in [0, 1)$ due to the normalization factor and the fact that $A \in [0, 1]^{n \times n}$. For a matrix A, we define the ℓ^1-norm of A to be $\|A\|_1 = \frac{1}{n^3} \sum_{i,j=1}^{n} |A|_{i,j}$.

The main result in [16] is the following,

Theorem 4.1.1. *For every symmetric matrix $A \in [0,1]^{n\times n}$, there exists a Robinsonian matrix R such that*

$$\|A - R\|_1 \leq 14\Gamma_1(A)^{1/3}.$$

Moreover, the matrix R can be computed in polynomial time.

First we note that there is a close relationship between the parameter $\Gamma_1(A)$ and the ℓ^1-distance between A and the best possible Robinsonian approximation of A, in fact Lemma 2.3 in [16] combined with theorem 4.1.1 states that

$$\frac{1}{4}\Gamma_1(A) \leq \min_{R \text{ is a } \mathcal{R}-matrix} \|A - R\|_1 \leq 14\Gamma_1(A)^{1/3}.$$

In the following we will present some algorithms that finds the matrix R in theorem 4.1.1. We will divide this section into two subsections, in the first one we will study binary matrices and the second one is concerned with general matrices.

4.1.1 Robinsonian approximation for binary matrices

We start this section with some definitions. For any cell (a,b) with $1 \leq a < b \leq n$ define the *upper right* (UR) and *lower left* (LL) around (a,b) to be the sets

$$\begin{aligned}\mathrm{UR}(a,b) &= \{(i,j) : i < a < b < j\},\\ \mathrm{LL}(a,b) &= \{(i,j) : a \leq i \leq j \leq b\}.\end{aligned}$$

Then let $1_{\mathrm{UR}}(a,b)$ denote the number of ones in the matrix A in the set $\mathrm{UR}(a,b)$ and similarly denote $0_{\mathrm{LL}}(a,b)$ to be the number of zeros in the lower left region around (a,b), mathematically we can write

$$\begin{aligned}1_{\mathrm{UR}}(a,b) &= |\mathrm{UR}(a,b) \cap \{(i,j) : A_{ij} = 1\}|,\\ 0_{\mathrm{LL}}(a,b) &= |\mathrm{LL}(a,b) \cap \{(i,j) : A_{ij} = 0\}|.\end{aligned}$$

Note that we can calculate both $1_{\mathrm{UR}}(a,b)$ and $0_{\mathrm{LL}}(a,b)$ using the following recursion equations

$$\begin{aligned}1_{\mathrm{UR}}(a,b) &= 1_{\mathrm{UR}}(a-1,b) + 1_{\mathrm{UR}}(a,b+1) - 1_{\mathrm{UR}}(a-1,b+1) + A_{a,b},\\ 0_{\mathrm{LL}}(a,b) &= 0_{\mathrm{LL}}(a+1,b) + 0_{\mathrm{LL}}(a,b-1) - 0_{\mathrm{LL}}(a+1,b-1) + 1 - A_{a,b},\end{aligned}$$

with the convention that $1_{\mathrm{UR}}(a,b) = 0_{\mathrm{LL}}(a,b) = 0$ when $a \in \{0, n+1\}$ or $b \in \{0, n+1\}$ or $a > b$.

The idea of algorithm 3 is to count $1_{\mathrm{UR}}(a,b)$ for each cell, and put 1's in the cells where $1_{\mathrm{UR}}(a,b)$ is large and 0's on the others. Here, large means bigger than a certain threshold t that we will choose later. This procedure will in fact create a Robinsonian matrix, because the region containing 1's in the output matrix R is convex around the diagonal, in the sense

Algorithm 3: Robinsonian approximation of binary matrices.

Input : A binary matrix $A \in \mathcal{A}_n$, threshold t
Output: binary R-matrix $R \in \mathcal{A}_n$

```
for i := 1,2,..., n do
    for j := i,i+1,...,n do
        if 1_UR(i, j) < t then
            R_ij = 0; R_ji = 0
        else
            R_ij = 1; R_ji = 1
        end
    end
end
return R
```

that, if a cell contain 1, then so are all other cells closer to the diagonal in the same row or column.

In [16] the authors proved that, for a binary matrix $A \in \mathcal{A}_n$, if R is the output of algorithm 3 with threshold $t = n^2\sqrt{2\Gamma_1(A)}$ the following inequality holds

$$\|A - R\|_1 \leq 2^{13/4}\Gamma_1(A)^{1/4} + \frac{2}{n}. \tag{4.1.2}$$

Note that this inequality is not the same as the one in theorem 4.1.1, but it will be improved in the next subsections using the preprocessing idea.

4.1.2 Robinsonian approximation for general matrices

Let $A \in \mathcal{A}_n$ be a matrix, the idea here is to decompose A into a convex combination of binary matrices, then apply algorithm 3 to each term of the sum. Denote $\text{range}(A) = \{A_{ij} : 1 \leq i \leq j \leq n\}$, then put $\text{range}(A) \cup \{0\} = \{s_0, s_1, \ldots, s_m\}$ where $0 = s_0 < s_1 < \cdots < s_m$, and define the matrices $A^{(k)}$ for each $k = 1, 2, \ldots, m$ to be

$$A^{(k)}_{ij} = \begin{cases} 1 \text{ if } A_{ij} \geq s_k, \\ 0 \text{ otherwise.} \end{cases}$$

Then we can write A as a convex combination of the binary matrices $A^{(k)}$ in the following manner

$$A = \sum_{k=1}^{m} (s_k - s_{k-1})A^{(k)}. \tag{4.1.3}$$

For a given matrix $A \in \mathcal{A}_n$, the authors of [16] showed that, for thresholds $t_k = n^2\sqrt{2\Gamma_1(A^{(k)})}$ for all $1 \leq k \leq m$, the output of algorithm 4 satisfy the following inequality

$$\|A - R\|_1 \leq 10\Gamma_1(A)^{1/4}(1 + \mathcal{O}(n^{-1/4})) \tag{4.1.4}$$

Algorithm 4: Robinsonian approximation of general matrices.

Input : A binary matrix $A \in \mathcal{A}_n$, thresholds $t_1, \ldots, t_m$
Output: binary R-matrix $R \in \mathcal{A}_n$

Compute range$(A) \cup \{0\}$ as an ordered list s.
$R := 0$
for *k := 1,2,..., m* **do**
    Let $R^{(k)}$ be the output of algorithm 3 with inputs $(A^{(k)}, t_k)$.
    $R := R + (s_k - s_{k-1})R^{(k)}$
end
return R

Note that while this inequality is not the same as in theorem 4.1.1, the authors in [16] proposed a preprocessing (algorithm 5) step for algorithm 3, which takes into input a binary matrix A and a certain threshold t, and consists of changing the upper right and lower left of some certain cells in A so that the updated matrix satisfies $1_{\mathrm{UR}}(a, b) < t$ or $0_{\mathrm{LL}}(a, b) < t$ for all cells (a, b).

Algorithm 5: Preprocessing step.

Input : A binary matrix $A \in \mathcal{A}_n$, threshold t
Output: Updated matrix A

for $i := 1, 2, \ldots, n$ **do**
    for $j := 1, 2, \ldots, n$ **do**
        if $1_{\mathrm{UR}}(i, j) \geq t$ *and* $0_{\mathrm{LL}}(i, j) \geq t$ **then**
            for *all cells* $(r, s) \in \mathrm{UR}(i, j)$ **do**
                $A_{rs} = 0$; $A_{sr=0}$
            end
            for *all cells* $(r, s) \in \mathrm{LL}(i, j)$ **do**
                $A_{rs} = 1$; $A_{sr=1}$
            end
        end
    end
end
return R

It turns out that, due to this final step theorem 4.1.1 holds. First, for a binary matrix A, apply algorithm algorithm 5 with threshold $t = \Gamma_1(A)^{2/3}n^2$ to get a matrix $\widehat{A}$ and then compute the output R of algorithm 3 with the same threshold t. Then it was shown that R satisfies the inequality in theorem 4.1.1. For a a general matrix $A \in \mathcal{A}_n$, the idea is the same using the convex combination decomposition in equation (4.1.3), and using the triangular inequality of the ℓ^1-norm and Jenssen's inequality to prove theorem 4.1.1.

4.2 Optimal rates of statistical seriation

In the paper [10] the authors considered the following seriation model; suppose we observe a matrix $Y \in \mathbb{R}^{n\times m}$ such that

$$Y = \Pi^* A^* + Z,$$

where A^* is a $\mathcal{Q}$-matrix, Π^* is a permutation matrix and Z is a sub-Gaussian noise matrix, with proxy $\sigma^2 > 0$. That is, Z satisfies $\mathbb{E}[Z] = 0$ and for any matrix $M \in \mathbb{R}^{n\times m}$ we have

$$\mathbb{E}\left[\exp\left(\mathrm{Tr}\left(Z^T M\right)\right)\right] \leq \exp\left(\frac{\sigma^2 \|M\|_F^2}{2}\right).$$

Then the goal of noisy seriation here is to estimate the unknown couple (Π^*, A^*) accurately. Note that the performance of an estimate $(\widehat{\Pi}, \widehat{A})$ is measured by the quadratic loss:

$$\frac{1}{nm}\left\|\widehat{\Pi}\widehat{A} - \Pi^* A^*\right\|_F^2.$$

Where $\|\cdot\|_F$ is the Frobenius norm. The paper focuses on the least square estimate (LS) defined by

$$(\widehat{\Pi}, \widehat{A}) \in \underset{(\Pi, A)\in \mathfrak{S}_n \times \mathcal{Q}_{m,n}}{\operatorname{argmin}} \|Y - \Pi A\|_F^2. \tag{4.2.1}$$

Here $\mathfrak{S}_n$ denote the set of $n \times n$ permutation matrices and $\mathcal{Q}_{m,n}$ denote the set of $m \times n$ $\mathcal{Q}$-matrices. Many upper bounds inequalities for the LS estimate in equation (4.2.1) were presented in [10], for the sake of this thesis we are going to show only one inequality, called the *adaptive oracle inequality*. Before stating the inequality, we denote for each matrix $A \in \mathcal{Q}_{n,m}$, $k(A_{\cdot,j}) = \mathrm{Card}(\{A_{1,j}, \ldots, A_{n,j}\})$ to be the number of distinct values in the j-th column. Then consider $K(A) = \sum_{j=1}^m k(A_{\cdot,j})$.

Theorem 4.2.1. *Let $A^* \in \mathcal{Q}_{n,m}$ and $Y = \Pi^* A^* + Z$ where Z is a sub-Gaussian noise with proxy σ^2. Then the LS estimate $(\widehat{\Pi}, \widehat{A})$ in equation (4.2.1) satisfies*

$$\frac{1}{nm}\left\|\widehat{\Pi}\widehat{A} - \Pi^* A^*\right\|_F^2 \leq A\sigma^2\left(\frac{K(A^*)}{nm}\log\left(\frac{enm}{K(A^*)}\right) + \frac{\log n}{m}\right).$$

with probability $1 - e^{-c(n+1)}$, $c > 0$ and $A > 0$ are two constants.

4.3 Convex relaxation for the seriation problem

In chapter 3 we presented a result from [12] which claims that the seriation problem is equivalent to the combinatorial 2-SUM problem. This result allows us to write seriation as a quadratic minimization problem over permutation matrices. The authors in [12] then produce convex relaxations for this last problem. It turns out that in practice the convex

relaxation is more robust to noise than the classical spectral or combinatorial methods presented in chapter 3. First, note that we might write the problem in (3.1.6) in the following way:

$$\begin{aligned} &\text{Minimize} \quad g^T \Pi^T L_A \Pi g \\ &\text{Subject to} \quad \Pi \in \mathfrak{S}_n. \end{aligned} \tag{4.3.1}$$

where $g = (1, 2, \ldots, n)^T$. This is an optimization problem over the set $\mathfrak{S}_n$ of permutation matrices, which is clearly not a convex problem. It is very known that the convex hull of $\mathfrak{S}_n$ is the set of doubly stochastic matrices, which is defined by

$$\mathcal{D}_n = \left\{ X \in \mathbb{R}^{n \times n} : X \geq 0, Xe = 1, X^T e = 1 \right\}, \tag{4.3.2}$$

where $e = (1, 1, \ldots, 1)^T$. We also have $\mathfrak{S}_n = \mathcal{D}_n \cap \mathcal{O}_n$ where $\mathcal{O}_n$ is the set of orthogonal matrices, i.e., a matrix is a permutation matrix if and only if it is both doubly stochastic and orthogonal. This facts allow us to write a convex relaxation of (4.3.1) by replacing $\mathfrak{S}_n$ with its convex hull $\mathcal{D}_n$, to get

$$\begin{aligned} &\text{Minimize} \quad g^T \Pi^T L_A \Pi g \\ &\text{Subject to} \quad \Pi \in \mathcal{D}_n. \end{aligned} \tag{4.3.3}$$

By symmetry, if a vector Πy minimizes (4.3.3), then the reverse vector also minimizes (4.3.3). We can add a linear constraint like $e_1^T \Pi g + 1 \leq e_n^T \Pi g$ to break symmetries. To improve the robustness to noise we consider the problem

$$\begin{aligned} &\text{Minimize} \quad \operatorname{Tr}\left(Y^T \Pi^T L_A \Pi Y\right)/p \\ &\text{Subject to} \quad \Pi \in \mathcal{D}_n, e_1^T \Pi g + 1 \leq e_n^T \Pi g. \end{aligned} \tag{4.3.4}$$

Where $Y \in \mathbb{R}^{n \times p}$ whose columns are small perturbations of $g = (1, 2, \ldots, n)^T$. Note that this optimization problem is only over the doubly stochastic matrices, we can add a penalty term to the objective function, so that the solution gets closer to orthogonal matrices. We can do this by noticing that a doubly stochastic of Frobenius norm $\sqrt{n}$ is necessarily orthogonal. For this purpose, the authors in [12] proposed the following problem

$$\begin{aligned} &\text{Minimize} \quad \frac{1}{p}\operatorname{Tr}\left(Y^T \Pi^T L_A \Pi Y\right) - \frac{\mu}{p}\left\|P\Pi\right\|_F^2 \\ &\text{Subject to} \quad \Pi \in \mathcal{D}_n, e_1^T \Pi g + 1 \leq e_n^T \Pi g. \end{aligned} \tag{4.3.5}$$

where $P = I_n - \frac{1}{n} ee^T$ and $\mu > 0$ is a real number so that $\mu \leq \lambda_2(L_A)\lambda_1(YY^T)$ to ensure that (4.3.5) is a convex problem. The quadratic programming relaxation above allows us to add constraints to the problem. For instance, in archeological applications, one may specify that observation i must appear before observation j, i.e. $\pi(i) < \pi(j)$. We can also impose other constraints such us $a \leq \pi(i) - \pi(j) \leq b$, which can be also written as $a \leq e_i^T \Pi g - e_j^T \Pi g \leq b$. More generally we can rewrite problem (4.3.5) with n_c additional linear constraints as follows

$$\begin{aligned} &\text{Minimize} \quad \frac{1}{p}\mathrm{Tr}\left(Y^T\Pi^T L_A \Pi Y\right) - \frac{\mu}{p}\left\|P\Pi\right\|_F^2 \\ &\text{Subject to} \quad \Pi \in \mathcal{D}_n, D^T\Pi g + \delta \leq 0. \end{aligned} \tag{4.3.6}$$

where D is a matrix of size $n \times (n_c + 1)$ and δ is a vector of size n_c. The matrix D and the vector δ contains the constrains to add to the optimization problem, the first column of D is equal to $e_1 - e_n$ and $\delta_1 = 1$ (to break symmetry). In [12], the authors implemented problem (4.3.6) for noisy archeological data and proved that, adding a reasonable number of constrains improve the robustness to noise of the seriation problem.

Chapter 5

Consistency of spectral seriation

5.1 Introduction

In this chapter we will present our contribution to the robustness to noise of the spectral seriation algorithm 1 presented in chapter 3. We will prove that for a general non-parametric family of similarity functions, that the spectral seriation algorithm is robust to noise.

Let P be a $N \times N$ pre-$\mathcal{R}$ matrix with coefficients in the interval $[0, 1]$, we will refer to the matrix P as the *model matrix*. As presented in chapter 3, let x be the Fiedler vector of P, and π be a permutation such that $x_{\pi(1)} \leq x_{\pi(2)} \leq \cdots \leq x_{\pi(N)}$, then this specific permutation π makes P^π a $\mathcal{R}$-matrix (see algorithm 1). Now, consider the random matrix $\widehat{P}$ defined as follows: for each $1 \leq i \leq j \leq N$ consider $\widehat{P}_{i,j} \sim \mathcal{B}\text{ernoulli}(P_{i,j})$, then put $\widehat{P}_{j,i} = \widehat{P}_{i,j}$ (that is $\widehat{P}$ is symmetric). We can see $\widehat{P}$ as the adjacency matrix of a graph with N vertices $\{1, 2, \ldots, N\}$, where we put an edge between i and j with probability $P_{i,j}$. Note that $\widehat{P}$ is not a pre-$\mathcal{R}$ matrix but very close to one. If we apply the spectral seriation algorithm 1 to the random matrix $\widehat{P}$ it will produce a permutation $\widehat{\pi}$. The question we are concerned with is how close $\widehat{\pi}$ to the true model permutation π? Without loss of generality we can assume that the true permutation π is the identity, that is the model matrix is a $\mathcal{R}$-matrix.

Figure 5.1 shows an example of a noisy pre-$\mathcal{R}$ matrix, which was simulated as follows; For $N = 100$, first choose a random permutation $\pi \in \mathcal{P}_N$. Then, consider the model matrix P of dimension $N \times N$, where $P_{i,j} = 1 - |i - j| / N$ for all $1 \leq i, j \leq N$. Next, construct the random matrix $\widehat{P}$ following a Bernoulli distribution (as described previously). Figure 5.1 shows the matrix $M = \widehat{P}^\pi$. Now, apply the spectral seriation algorithm 1 for the matrix M to get a permutation $\widehat{\pi}$. Figure 5.2 shows $M^{\widehat{\pi}}$. Note that the figure on the right, makes the structure of the matrix almost a $\mathcal{R}$-matrix.

The goal of this chapter is the following; we consider an infinite family of model matrices $(P^{(N)})_N$, each matrix $P^{(N)}$ is a $\mathcal{R}$-matrix of size $N \times N$, and these matrices satisfy certain assumptions that will be discussed later. The output of the spectral seriation algorithm 1 is $\pi^{(N)} = Id_N$ (since $P^{(N)}$ is a $\mathcal{R}$-matrix). Then we consider $\hat{\pi}^{(N)}$ to be the output of algorithm 1 for the matrix $(\widehat{P}^{(N)})_N$ (which is a random matrix constructed as explained earlier). In this work, we will show that $F(\hat{\pi}^{(N)}, \pi^{(N)}) = o(N^2)$ with high probability, where

F is the Spearman's footrule distance. That is indeed, in these settings, the spectral seriation algorithm is robust to noise.

This chapter is organized as follows: In section 5.2, we start by some definitions and notations as well as the statement of the problem we are concerned with. In section 5.3, we prove, under some hypothesis on the graphon, several important properties on the Fiedler function, which is the eigenfunction of the Laplacian operator. Next, in section 5.4, we prove the convergence of the Fiedler function using results from probability theory such as Bernstein inequality. These sections will allow us to derive bounds on the permutations, which will be detailed in section 5.5.

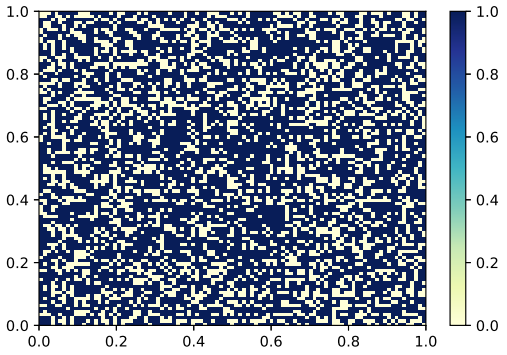


Figure 5.1: Example of a noisy pre-$\mathcal{R}$ matrix.

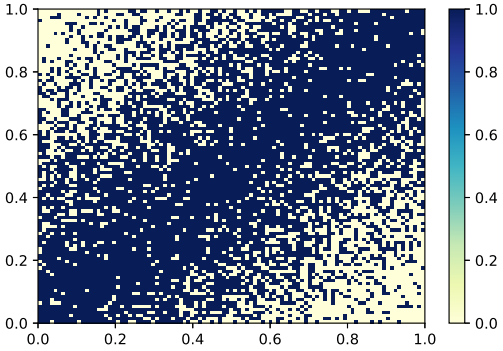


Figure 5.2: The noisy matrix in Figure 5.1 after an appropriate permutation.

5.2 Notations and Problem Statement

A *graphon* is a measurable function $w : [0,1]^2 \to [0,1]$ which is symmetric (i.e., $w(x,y) = w(y,x)$). Graphons are used in the context of seriation, see for example [4]. For more results on graphons we suggest [22]. Graphons shall be important in constructing the model matrix P.

We say that a graphon w is a *$\mathcal{R}$-graphon* or *Robinsonian graphon* if the following property holds: for all $x, y, z \in [0,1]$ we have

$$\begin{aligned} y < z < x &\Rightarrow w(x,y) \leq w(x,z), \\ x < y < z &\Rightarrow w(x,y) \geq w(x,z). \end{aligned} \tag{5.2.1}$$

In [4], the authors refer to such graphons as, graphons with a linear embedding.

Here we are given an infinite family of matrices $(P^{(N)})_{N\in\mathbb{N}}$ where each $P^{(N)}$ is a $\mathcal{R}$-matrix with $P^{(N)}_{i,j} = w(i/N, j/N)$ for all $1 \leq i, j \leq N$, for some $\mathcal{R}$-graphon w. In [26] the authors worked with the model $w(x,y) = \mathbb{I}_{\{|x-y|\leq 1/2\}}$, and in [11] the authors considered the

function $w(x,y) = 1 - |x-y|$. In these two papers the question of how close $\widehat{\pi}$ to π was answered, but for only these two specific models. Their method relies on the calculation of the spectrum of the model matrix M, which is impossible in general. Here we consider a different approach using results from spectral theory of linear operators, our method can be applied to a general class of graphons.

From here on, we denote by $L^2([0,1])$ the Hilbert function space with inner product

$$\langle f, g\rangle := \int_0^1 f(x)g(x)dx,$$

for all $f, g \in L^2([0,1])$, and norm $\|f\| := \sqrt{\langle f, f\rangle}$. The elements of $L^2([0,1])$ are the equivalence classes of Lebesgue square-integrable functions $f : [0,1] \to \mathbb{R}$, that is, we identify two functions f and g with each other if they differ only on a set of measure zero (i.e., $f \equiv g \Leftrightarrow \|f - g\| = 0$). Let $e \in L^2([0,1])$ denote the constant function: $e(x) = 1$ for all $x \in [0,1]$. We use blackboard bold symbols, such as $\mathbb{T}$, to denote linear operators acting on $L^2([0,1])$. The induced (operator) norm is defined as

$$|||\mathbb{T}||| := \sup_{f \in L^2([0,1]) \text{ s.t. } \|f\|=1} \|\mathbb{T}f\|.$$

For a bounded self-adjoint operator $\mathbb{T}$ in $L^2([0,1])$ we denote by $\sigma(\mathbb{T})$ its spectrum, note that the spectrum is always non-empty, closed and bounded [24]. We define the *discrete spectrum* $\sigma_{\mathrm{d}}(\mathbb{T})$ to be the part of $\sigma(\mathbb{T})$ which consists of all isolated eigenvalues with finite algebraic multiplicity, and the *essential spectrum* $\sigma_{\mathrm{ess}}(\mathbb{T}) = \sigma(\mathbb{T}) \setminus \sigma_{\mathrm{d}}(\mathbb{T})$. The essential spectrum is always closed, and the discrete spectrum can only have accumulation points on the boundary to the essential spectrum, for more results on the spectrum of linear operator we suggest [24]. Denote by $r(\mathbb{T})$ the spectral radius of $\mathbb{T}$,

$$r(\mathbb{T}) = \sup\{|\lambda| : \lambda \in \sigma(\mathbb{T})\}$$

Note that $r(\mathbb{T}) \leq |||\mathbb{T}|||$. In chapter 3 we introduced the unnormalized Laplacian of a symmetric matrix A as a matrix L defined by $L = D_A - A$ where D_A is the degree matrix. We know that for all $x \in \mathbb{R}^n$,

$$Lx = \sum_{j=1}^{n} a_{ij}(x_i - x_j)$$

We can generalize this definition to graphons, by defining the *Graphon-Laplacian operator* $\mathbb{L}$. For a given graphon w let $\mathbb{L}$ be the operator defined as, for all $f \in L^2([0,1])$ and $x \in [0,1]$,

$$\begin{aligned}(\mathbb{L}f)(x) &:= \int_0^1 w(x,y)(f(x) - f(y))dy \\ &= f(x)d(x) - \int_0^1 w(x,y)f(y)dy\end{aligned} \tag{5.2.2}$$

where

$$d(x) = \int_0^1 w(x,y)dy$$

is the degree of x. Note that the operator $(\mathbb{W}f)(x) := \int_0^1 w(x,y)f(y)dy$ is a Hilbert-Schmidt operator (see e.g. [24]), thus $\mathbb{W}$ is compact. Nevertheless, $\mathbb{L}$ is not compact since it is the sum of a compact operator and a multiplication operator (see e.g. [28]). It is well known that compact perturbations do not affect the essential spectrum (see e.g., [15]), therefore the essential spectrum of $\mathbb{L}$ is the range of the degree function d, i.e., $\sigma_{\text{ess}}(\mathbb{L}) = rg(d)$ (see e.g., [28]).

Theorem 5.2.1. *Let w be a graphon, and $\mathbb{L}$ be the associated graphon-laplacian operator. Then $\lambda = 0$ is the smallest eigenvalue of $\mathbb{L}$. Moreover if $w(x,y) > 0$ for almost all $(x,y) \in [0,1]^2$, then $\lambda = 0$ is a simple eigenvalue.*

Proof. From the definition in equation (5.2.2), we have $\mathbb{L}e = 0$ hence $\lambda = 0$ is an eigenvalue of $\mathbb{L}$. It is the smallest eigenvalue, since:

$$\langle \mathbb{L}f, f\rangle = \frac{1}{2}\int_0^1 \int_0^1 w(x,y)(f(x) - f(y))^2 dxdy \geq 0. \tag{5.2.3}$$

Now, if $w(x,y) > 0$ for almost all $(x,y) \in [0,1]^2$, then if $\mathbb{L}f = 0$ for some $f \in L^2([0,1])$, equation (5.2.3) implies that $f(x) = f(y)$ for almost all $(x,y) \in [0,1]^2$, i.e., $f \in \text{Vect}(e)$, that is $\lambda = 0$ is simple. ∎

Denote $\lambda_1 = 0 \leq \lambda_2 \leq \cdots$ the eigenvalues of $\mathbb{L}$ in increasing order. Similar to chapter 3, we call the *Fiedler value* the smallest non-zero eigenvalue (i.e., λ_2) and any corresponding eigenfunction ϕ a *Fiedler function*. In the next section, we will prove that under some specific conditions on w, the Fiedler value λ_2 is simple and the corresponding Fiedler vector ϕ is monotone.

For each integer N consider the partition $(\mathcal{I}_i^N)_{i=1,\ldots,N}$ of the interval $[0,1]$ defined by $\mathcal{I}_i^N = [(i-1)/N, i/N)$ for $i = 1,\ldots,N-1$ and $\mathcal{I}_N^N = [(N-1)/N, 1]$. For a given graphon w, as presented in the introduction of this chapter, we define the model matrix $P^{(N)} \in [0,1]^N$ by $P_{ij}^{(N)} := w(i/N, j/N)$ for all $i,j \in \{1,2,\ldots,N\}$. We also define the sampled graphon w_N by

$$w_N(x,y) := \sum_{i=1}^N \sum_{j=1}^N P_{ij}^{(N)} 1_{\mathcal{I}_i^N}(x) 1_{\mathcal{I}_j^N}(y)$$

for all $(x,y) \in [0,1]^2$. Furthermore, let $\widehat{P}^{(N)}$ denote the adjacency matrix of a symmetric (random) graph obtained by taking N isolated vertices $i \in \{1,\ldots,N\}$ and adding undirected edges between vertices i and j at random with probability $\rho_N P_{ij}^{(N)}$ (i.e., $\widehat{P}_{ij}^{(N)} \sim$

$\mathcal{B}$ernoulli$(\rho_N P_{ij}^{(N)})$) for all $i > j$. Where $\rho_N \in (0,1]$ is a constant regulating the sparsity of the graph. The associated random graphon $\widehat{w}_N$ is defined by:

$$\widehat{w}_N(x,y) := \sum_{i=1}^{N}\sum_{j=1}^{N} \widehat{P}_{ij}^{(N)} 1_{\mathcal{I}_i^N}(x) 1_{\mathcal{I}_j^N}(y)$$

Let $\mathbb{L}_N$ and $\widehat{\mathbb{L}}_N$ denote the Laplacian operators associated to graphons w_N and $\widehat{w}_N$ respectively, as defined in equation (5.2.2).

Our method relies on the following conditions on the graphon:

Assumption 5.2.2. *$w(x,y) > 0$ for almost all $(x,y) \in [0,1]^2$.*

Assumption 5.2.3. *Assume that for all $x \in [0,1]$, $w(x,\cdot)$ is continuous almost everywhere. Moreover, assume that the partial derivative $\partial w(x,\cdot)/\partial x$ exists and is non zero and bounded almost everywhere.*

Assumption 5.2.4. *Let d be the degree function associated with w, assume that the set $N := \{x \in [0,1] : d'(x) = 0\}$ is countable.*

Assumption 5.2.5. *Let*

$$\mathcal{H} = \left\{ f \in L^2([0,1]) : \int_0^1 f(t)dt = 0 \text{ and } \int_0^1 f^2(t)dt = 1 \right\},$$

and $\mathbb{L}$ be the graphon-Laplacian operator associated with w. Assume that

$$\inf_{f\in\mathcal{H}} \langle \mathbb{L}f, f\rangle < \min_{x\in[0,1]} d(x).$$

Note that the inequality in assumption 5.2.5 is in fact ensuring that $\lambda_2 < \min_{x\in[0,1]} .d(x)$.

Assumption 5.2.6 (Lipschitz graphon)**.** *There exists a constant $K > 0$ such that*

$$|w(x,y) - w(x',y')| \le K(|x-x'| + |y-y'|)$$

for all pairs $(x,y) \in [0,1]^2, (x',y') \in [0,1]^2$.

In the next section we will study the properties of the spectrum of $\mathbb{L}$ under these assumptions. Before starting the next section, we present an important theorem that shows that these assumptions are satisfied for a large family of graphons.

Theorem 5.2.7. *Let $R : [0,1] \to [0,1]$ be a $\mathcal{C}^1$ function such that $R'(x) < 0$ for all $x \in (0,1)$. Define the graphon $w(x,y) = R(|x-y|)$ for all $x, y \in [0,1]$, then w is a $\mathcal{R}$-graphon satisfying all the above assumptions.*

Proof. To see that w is a $\mathcal{R}$-graphon, consider $x, y, z \in [0,1]$, if $y < z < x$ then $|x-y| > |x-z|$ therefore $R(|x-y|) \leq R(|x-z|)$ (since R is non-increasing), thus $w(x,y) \leq w(x,z)$. By the same argument, if $x < y < z$, then $w(x,y) \geq w(x,z)$. Assumptions 5.2.2 and 5.2.3 follows immediately from the fact that R is differentiable and satisfies $R'(x) < 0$ for almost all $x \in [0,1]$. Next, note that

$$\begin{aligned} d(x) &= \int_0^1 R(|x-y|)dy \\ &= \int_0^x R(x-y)dy + \int_x^1 R(y-x)dy \\ &= \int_0^x R(z)dz + \int_0^{1-x} R(z)dz \end{aligned}$$

by a change of variable. Then $d'(x) = R(x) - R(1-x)$. Consequently $d'(x) = 0$ if and only if $x = 1/2$ (again because $R'(x) < 0$ for all x). This proves assumption 5.2.4. Now to prove that w is a Lipschitz graphon, let K such that $|R(x) - R(x')| \leq K|x - x'|$ for all $x, x' \in [0,1]$. Then for all $(x,y), (x',y') \in [0,1]^2$, by the triangular inequality we have

$$\begin{aligned} |w(x,y) - w(x',y')| &= |R(|x-y|) - R(|x'-y'|)| \\ &\leq K\,||x-y| - |x'-y'|| \\ &\leq K\,(|x-x'| + |y-y'|) \end{aligned}$$

The last thing to prove is assumption 5.2.5 which is the hardest part of this proof. First we note that $\min_{x\in[0,1]} d(x) = d(0) = \int_0^1 R(x)dx$ and by equation 5.2.3, the inequality of assumption 5.2.5 is equivalent to

$$\inf_{f\in\mathcal{H}} \left(\int_0^1 \int_0^1 R(|x-y|)(f(x)-f(y))^2 dxdy \right) < 2\int_0^1 R(x)dx$$

Let $f \in \mathcal{H}$, note that

$$\int_0^1 \int_0^1 R(|x-y|)(f(x)-f(y))^2\,dx\,dy - 2\int_0^1 R(x)dx = 2D_f(R),$$

where

$$D_f(R) := \int_0^1 \int_x^1 R(y-x)(f(x)-f(y))^2 dydx - \int_0^1 R(z)dz.$$

So, it is enough to find, a function $f \in \mathcal{H}$ such that $D_f(R) < 0$. Take any $t \in (0, 1/2)$, and for all $x \in [0,1]$ let

$$f(x) := f_t(x) := \tfrac{1}{\sqrt{2t}}(I\{x < t\} - I\{x > 1-t\}),$$

where I is the indicator. Then $f \in \mathcal{H}$. Also,

$$2t(f(x)-f(y))^2 = I\{x<t\}I\{y \geq t\} + I\{x \leq 1-t\}I\{y > 1-t\} + 2I\{x<t\}I\{y \geq 1-t\} \tag{5.2.4}$$

if $0 < x < y < 1$, whence, letting $t \downarrow 0$, we have

$$\int_0^1 \int_x^1 R(y-x)(f(x)-f(y))^2 dydx = \frac{J_1 + J_2 + 2J_3}{2t},$$

where J_1, J_2 and J_3 are the terms in equation (5.2.4). Moreover, we have:

$$\begin{aligned}
J_1 &:= \int_0^t \int_t^1 R(y-x)dydx \\
&= \int_0^t \int_{t-x}^{1-x} R(u)dudx \\
&= t\int_0^1 R(z)dz - \int_0^t \Big(\int_0^{t-x} R(z)dz + \int_{1-x}^1 R(z)dz \Big) dx \\
&= t\int_0^1 R(z)dz - \int_0^t \Big(\int_0^{t-x} [R(0)+o_t(1)] + \int_{1-x}^1 [R(1)+o_t(1)] \Big) dx \\
&= t\int_0^1 R(z)dz - \frac{t^2}{2}\left[R(0)+R(1)+o_t(1)\right];
\end{aligned}$$

here $o_t(1)$ is a quantity that goes to 0 as t goes to 0. Similarly,

$$\begin{aligned}
J_2 &:= \int_{1-t}^1 \int_0^{1-t} R(y-x)dxdy \\
&= t\int_0^1 R(z)dz - \frac{t^2}{2}[R(0)+R(1)+o_t(1)];
\end{aligned}$$

and

$$\begin{aligned}
J_3 &:= \int_0^t \int_{1-t}^1 R(y-x)dydx \\
&= t^2[R(1)+o_t(1)].
\end{aligned}$$

Collecting all the pieces, we have

$$D_f(R) = \frac{J_1+J_2+2J_3}{2t} - \int_0^1 R(z)dz = [R(1)-R(0)+o_t(1)]t/2 < 0$$

for small enough $t > 0$ (because $R(1) < R(0)$). This finishes the proof of Theorem 5.2.7 ∎

5.3 Properties of the Fiedler value and the Fiedler function

In the following we consider a graphon w satisfying Assumptions 5.2.2, 5.2.3 and 5.2.4.

Theorem 5.3.1. *Let λ be an eigenvalue of $\mathbb{L}$ and f any associated eigenfunction. Define $D_\lambda := d^{-1}\{\lambda\}$ then the set D_λ is discrete and f is $\mathcal{C}^1$ in D_λ^c.*

In particular if $\lambda \notin rg(d)$ then f is $\mathcal{C}^1$ in $[0,1]$.

Proof. First we show that under Assumption 5.2.4 the set D_λ is discrete. Let $N = \{x \in [0,1] : d'(x) = 0\}$ as in assumption 5.2.4, since N is discrete it is enough to show that $D' = D_\lambda \setminus N$ is discrete. Let $x \in D'$ then $d(x) = \lambda$ and $d'(x) \neq 0$. Moreover, $d(x+h) - \lambda = d(x+h) - d(x) = d'(x)h + o(h)$ hence $d(y) \neq \lambda$ for all $y \neq x$ in some neighborhood U_x of x. Consequently every point $x \in D'$ has a neighborhood U_x such that $D' \cap U_x = \{x\}$ as required.

Now take $x \in D_\lambda^c$ and δ such that $x + \delta \in D_\lambda^c$. By definition of the operator $\mathbb{L}$ (equation (5.2.2)). We have:

$$f(x)d(x) - \int_0^1 f(y)w(x,y)dy = \lambda f(x)$$

Plugging x and $x+\delta$ in the previous equation and subtracting the two yields to:

$$\begin{aligned}\frac{f(x+\delta) - f(x)}{\delta} &= \frac{1}{d(x+\delta) - \lambda}\left(\int_0^1 (f(y) - f(x))\left(\frac{w(x+\delta,y) - w(x,y)}{\delta}\right)dy\right) \\ &\xrightarrow[\delta\to 0]{} \frac{1}{d(x) - \lambda}\left(\int_0^1 (f(y) - f(x))\frac{\partial w(x,y)}{\partial x}\right)dy\end{aligned}$$

for the integral convergence, we applied the dominated convergence theorem, note that by Assumption 5.2.3 we have that $(w(x+\delta,y) - w(x,y))/\delta$ is bounded for small δ. Therefore f is differentiable in D_λ^c and for all $x \in D_\lambda^c$

$$f'(x) = \frac{1}{d(x) - \lambda}\left(\int_0^1 (f(y) - f(x))\frac{\partial w(x,y)}{\partial x}\right)dy. \tag{5.3.1}$$

By a similar argument we can show that $f'(x+\delta) \to f'(x)$ for $\delta \to 0$, consequently f is $\mathcal{C}^1$ in D_λ^c.

∎

We are now ready to introduce the two main theorems of this section:

Theorem 1. *Let w be a $\mathcal{R}$-graphon satisfying assumptions 5.2.2, 5.2.3, 5.2.4 and 5.2.5. Then there exists a Fiedler function $\phi \in \mathcal{C}^1([0,1])$ such that $\phi'(x) > 0$ for all $x \in [0,1]$.*

Theorem 2. *Under the same assumptions as in Theorem 1, the Fiedler value λ_2 is simple. That is if $\mathbb{L}\varphi = \lambda_2\varphi$, then $\varphi \in \mathrm{Vect}(\phi)$, where ϕ is the Fiedler function defined in Theorem 1.*

Corollary 5.3.2. *There exists $L > 0$ such that $|\phi(x) - \phi(y)| \geq L|x-y|$ for all $(x,y) \in [0,1]^2$.*

Proof of Corollary 5.3.2. By Theorem 1, ϕ is a $\mathcal{C}^1$ function. It is enough to take $L = \inf_{x\in[0,1]} \phi'(x) > 0$ and apply the mean value theorem. ∎

The proof of Theorem 1 relies strongly on the following two Lemmas:

Lemma 5.3.3. *Let $\mathbb{M}$ be the linear operator defined by*

$$(\mathbb{M}f)(x) = d(x)f(x) - \int_0^1 \int_0^1 \left(f(z) \frac{\partial w(x,y)}{\partial x} \mathbf{1}_{\{x\le z\le y\}} \right) dzdy \tag{5.3.2}$$

then the eigenvalues of $\mathbb{M}$ are $\lambda_2 \le \lambda_3 \le \cdots$ (i.e., $\sigma(\mathbb{M}) = \sigma(\mathbb{L}) \setminus \{0\}$). Moreover for all $\lambda \neq 0 : (\lambda, f)$ is an eigenvalue-eigenfunction pair of $\mathbb{L}$ if and only if (λ, f') is an eigenvalue-eigenfunction pair of $\mathbb{M}$.

Lemma 5.3.4. *There exists $f^* \in L^2([0,1])$ such that $\mathbb{M}f^* = \lambda_2 f^*$ and $f^*(x) > 0$ for all $x \in [0,1]$.*

Proof of Lemma 5.3.3. Let (λ, f) be an eigenvalue-eigenfunction pair of $\mathbb{L}$ where $\lambda \neq 0$. Then by Theorem 5.3.1 f is differentiable at all $x \in D_\lambda^c$, therefore

$$\begin{aligned}
(\mathbb{L}f)(x) = \lambda f(x) &\Rightarrow \lambda f(x) = f(x)d(x) - \int_0^1 f(y)w(x,y)dy \\
&\Rightarrow \lambda f'(x) = f'(x)d(x) + f(x)d'(x) - \int_0^1 f(y)\frac{\partial w(x,y)}{\partial x}dy \\
&\Rightarrow \lambda f'(x) = f'(x)d(x) - \int_0^1 (f(y) - f(x))\frac{\partial w(x,y)}{\partial x}dy \\
&\Rightarrow \lambda f'(x) = f'(x)d(x) - \int_0^1 \int_0^1 \left(f'(z)\frac{\partial w(x,y)}{\partial x}\mathbf{1}_{\{x\le z\le y\}} \right) dzdy \\
&\Rightarrow \lambda f'(x) = (\mathbb{M}f')(x).
\end{aligned}$$

Conversly, let (λ, g) be an eigenvalue-eigenfunction pair of $\mathbb{M}$ then for all $x \in D_\lambda^c$ it is straightforward that

$$g(x) = \frac{1}{d(x) - \lambda} \int_0^1 \int_0^1 \left(\frac{\partial w(x,y)}{\partial x} g(z) \mathbf{1}_{\{x\le z\le y\}} \right) dzdy.$$

Applying the dominated convergence theorem with Assumption 5.2.3, g is continuous in D_λ^c. Define $f(x) = \int_0^1 g(t)dt$ for all $x \in [0,1]$, then for $x \in D_\lambda^c : f'(x) = g(x)$ and

$$(\mathbb{M}g)(x) = \lambda g(x) \Rightarrow d(x)f'(x) + \int_0^1 (f(x) - f(y))\frac{\partial w(x,y)}{\partial x}dy = \lambda f'(x)$$

$$\Rightarrow \int_0^1 \left(\frac{\partial (f(x) - f(y))}{\partial x} w(x,y) + (f(x) - f(y)) \frac{\partial w(x,y)}{\partial x} \right) dy = \lambda f'(x)$$
$$\Rightarrow \int_0^1 \frac{\partial}{\partial x} \left((f(x) - f(y)) w(x,y) \right) dy = \lambda f'(x)$$
$$\Rightarrow \int_0^1 (f(x) - f(y)) w(x,y) dy = \lambda f(x)$$
$$\Rightarrow (\mathbb{L}f)(x) = \lambda f(x).$$

As required. Finally, if $\mathbb{M}g = 0$, then $\mathbb{L}f = 0$ thus $f \in \text{Vect}(e)$ (since $\lambda_1 = 0$ is simple by Theorem 5.2.1) and therefore $g = f' = 0$. That is $\lambda = 0 \notin \sigma(\mathbb{M})$. We conclude that $\sigma(\mathbb{M}) = \sigma(\mathbb{L}) \setminus \{0\}$. ∎

Proof of Lemma 5.3.4. Let $\alpha > 1 + r(\mathbb{M})$ where $r(\mathbb{M})$ is the spectral radius of $\mathbb{M}$. Consider the operator $\mathbb{M}_\alpha = \alpha \mathbb{I} - \mathbb{M}$ with $\mathbb{I}$ the identity operator. It is obvious that $\sigma(\mathbb{M}_\alpha) = \alpha - \sigma(\mathbb{M})$ hence $r(\mathbb{M}_\alpha) > 0$. We claim, $\mathbb{M}_\alpha$ is positive with respect to the cone

$$K = \left\{ f \in L^2([0,1]) : f(x) \geq 0 \text{ for all } x \in [0,1] \right\}.$$

To see this, let $f \in K$ and $x \in [0,1]$. We have:

$$(\mathbb{M}_\alpha f)(x) = (\alpha - d(x)) f(x) + \int_0^1 \left(\int_x^y f(z) dz \right) \frac{\partial w(x,y)}{\partial x} dy$$

The first term $(\alpha - d(x)) f(x)$ is non-negative since $\alpha > 1 \geq d(x)$. For the second term, if $y > x$ then $\int_x^y f(z) dz \geq 0$ and $\frac{\partial w(x,y)}{\partial x} \geq 0$ (because w is $\mathcal{R}$-graphon). Similarly if $y < x$ both are negative, hence $\left(\int_x^y f(z) dz \right) \frac{\partial w(x,y)}{\partial x} \geq 0$ for all y. Therefore $\mathbb{M}_\alpha f \in K$, as required. Consequently $\mathbb{M}_\alpha$ is a positive, bounded and self-adjoint linear operator, with $\sigma_{\text{ess}}(\mathbb{M}_\alpha) = \alpha - rg(d)$ (because $\sigma_{\text{ess}}(\mathbb{L}) = rg(d)$). By Theorem 1 from [9], there exists $f^* \in K \setminus \{0\}$, such that $\mathbb{M}_\alpha f^* = r(\mathbb{M}_\alpha) f^*$.

Since $r(\mathbb{M}_\alpha)$ is the largest eigenvalue of $\mathbb{M}_\alpha$, and $\sigma(\mathbb{M}_\alpha) = \alpha - \sigma(\mathbb{M})$, then $\alpha - r(\mathbb{M}_\alpha)$ is the smallest eigenvalue of $\mathbb{M}$. Additionally, by Lemma 5.3.3 $\sigma(\mathbb{M}) = \sigma(\mathbb{L}) \setminus \{0\}$, thus the Fiedler value is $\lambda_2 = \alpha - r(\mathbb{M}_\alpha)$, and $\mathbb{M}f^* = \lambda_2 f^*$. Now, if $f^*(a) = 0$ for some a then $(\mathbb{M}f^*)(a) = \lambda_2 f^*(a) = 0$ which yields

$$\int_0^1 \left(\int_a^y f^*(z) dz \right) \frac{\partial w(a,y)}{\partial x} dy = 0$$

thus for almost all $y \in [0,1]$

$$\left(\int_a^y f^*(z) dz \right) \frac{\partial w(a,y)}{\partial x} = 0$$

by assumption 5.2.3 $\frac{\partial w(a,y)}{\partial x} \neq 0$ a.e., thus $\int_a^y f^*(z)dz = 0$ a.e., that is $f^* = 0$ in $L^2([0,1])$, which is a contradiction. Finally $f^*(x) > 0$ for all $x \in [0,1]$ as required. ∎

Proof of Theorem 1. First by assumption 5.2.5, $\lambda_2 \notin rg(d)$ because $\lambda_2 < \min_{x\in[0,1]} d(x)$. Thus by Theorem 5.3.1 all eigenfunctions of $\mathbb{L}$ associated with λ_2 are $\mathcal{C}^1([0,1])$. Let f^* as in Lemma 5.3.4, and define $\phi(x) = \int_0^x f^*(t)\mathrm{d}t$ for all $x \in [0,1]$. By Lemma 5.3.3 we know that $\mathbb{L}\phi = \lambda_2\phi$, thus $\phi \in \mathcal{C}^1([0,1])$ is a Fiedler function with $\phi'(x) = f^*(x) > 0$ for all $x \in [0,1]$. ∎

Finally, we present the proof of Theorem 2

Proof of Theorem 2. Let ϕ as in Theorem 1 and φ any other Fiedler function such that $\varphi \notin \mathrm{Vect}(\phi)$. Since $\lambda_2 \notin rg(d)$ then $\varphi \in \mathcal{C}^1([0,1])$ as well by theorem 5.3.1. Let a be such that

$$\frac{\varphi'(a)}{\phi'(a)} = \min_{x\in[0,1]} \left(\frac{\varphi'(x)}{\phi'(x)}\right).$$

Put $\psi = \phi'(a)\varphi - \varphi'(a)\phi$. Then ψ is an other Fiedler function satisfying $\psi'(x) \geq 0$ for all $x \in [0,1]$ and $\psi'(a) = 0$. But this implies $\mathbb{M}\psi'(a) = \lambda_2\psi'(a) = 0$ therefore $\psi = 0$ in $L^2([0,1])$ as in the proof of Lemma 5.3.4, which is a contradiction. ∎

5.4 Convergence of the Fiedler eigenfunction

In this section we will show that the Fiedler function of the operators $\mathbb{L}_N$ and $\widehat{\mathbb{L}}_N$ converge to the Fiedler function of the limiting operator, $\mathbb{L}$. Figure 5.3 shows how close the Fiedler vector of a model matrix P to the random matrix $\widehat{P}$.

5.4.1 Convergence of Laplacian operator

Theorem 5.4.1 (Convergence of graphon Laplacian)**.** *For a graphon w fulfilling Assumption 5.2.6 we have:*

$$|||\mathbb{L}_N - \mathbb{L}||| \leq \frac{4K}{N} \tag{5.4.1}$$

Moreover, for all $\varepsilon \in (0, 1/2)$ and $N \in \mathbb{N}$ large enough as in Lemma 5.4.2

$$\left|\left|\left|\rho_N^{-1}\widehat{\mathbb{L}}_N - \mathbb{L}\right|\right|\right| \leq 4\sqrt{\frac{\log(2N/\varepsilon)}{\rho_N N}} + \frac{4K}{N} \tag{5.4.2}$$

with probability at least $1 - 2\varepsilon$.

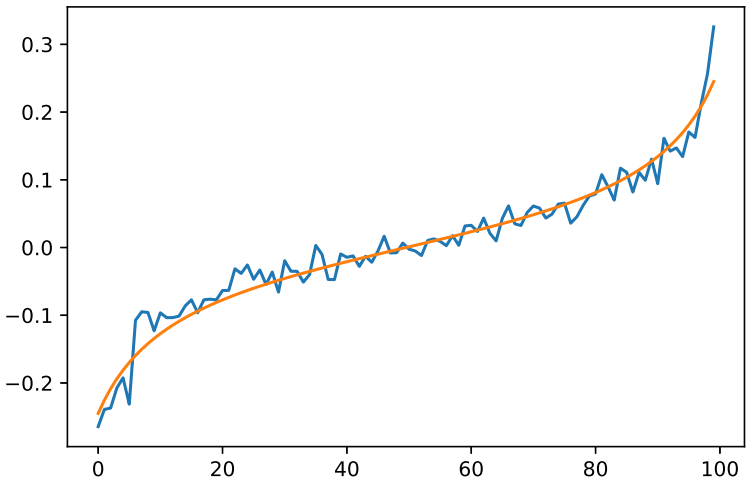


Figure 5.3: Convergence of the Fiedler function.

Lemma 5.4.2. *Let $\epsilon \in (0,1)$, if N is such that*

$$d_{\text{MAX}} := \max_{x\in[0,1]} d(x) \geq \frac{4\log(2N/\varepsilon)}{9\rho_N N} + \frac{2K}{N}$$

then the maximum expected degree of $P^{(N)}$

$$d^{(N)}_{\text{MAX}} := \max_{i=1}^{N} \left(\sum_{j=1}^{N} P^{(N)}_{ij} \right) \geq \frac{4\log(2N/\varepsilon)}{9\rho_N}$$

Proof of Lemma 5.4.2. Let $x \in \mathcal{I}^N_i$ for some $i = 1, 2, \ldots, N$ then

$$\begin{aligned} |d(x) - d_N(x)| &= \left| \int_0^1 (w(x,y) - w_N(x,y))\, dy \right| \\ &\leq \sum_{j=1}^{N} \int_{\mathcal{I}^N_j} \left| w(x,y) - w(\frac{i}{N}, \frac{j}{N}) \right| dy \\ &\leq \frac{2K}{N} \end{aligned}$$

where the inequality from line two to line three follows from Assumption 5.2.6. Therefore $d_N(x) \geq d(x) - \frac{2K}{N}$, so

$$\max_{x\in[0,1]} d_N(x) \geq d_{\text{MAX}} - \frac{2K}{N}$$

$$\geq \frac{4\log(2N/\varepsilon)}{9\rho_N N}$$

on the other hand,

$$\max_{i=1}^{N}\left(\sum_{j=1}^{N} P_{ij}^{(N)}\right) = N \max_{x\in[0,1]}\left(d_N(x)\right)$$
$$\geq \frac{4\log(2N/\varepsilon)}{9\rho_N}$$

∎

Proof of Theorem 5.4.1. We start by proving the first inequality

Proof of Inequality (5.4.1). Let $f \in L^2([0,1])$ such that $\|f\| = 1$,

$$\begin{aligned}
\|\mathbb{L}_N f - \mathbb{L} f\|^2 &= \int_0^1 \left(\mathbb{L}_N f - \mathbb{L} f\right)^2(x)dx \\
&= \int_0^1 \left(\int_0^1 (f(x)-f(y))(w_N(x,y)-w(x,y))dy\right)^2 dx \\
&\leq \int_0^1 \left(\int_0^1 (f(x)-f(y))^2 dy\right)\left(\int_0^1 (w_N(x,y)-w(x,y))^2 dy\right) dx \\
&= \sum_{i=1}^{N}\int_{\mathcal{I}_i^N} \left(\int_0^1 (f(x)-f(y))^2 dy\right)\left(\int_0^1 (w_N(x,y)-w(x,y))^2 dy\right) dx \quad (5.4.3)
\end{aligned}$$

where from line two to three we used Cauchy-Schwarz inequality. Let $x \in \mathcal{I}_i^N$ then

$$\begin{aligned}
\int_0^1 (w_N(x,y)-w(x,y))^2 dy &= \sum_{j=1}^{N}\int_{\mathcal{I}_j^N} (w(\frac{i}{N},\frac{j}{N})-w(x,y))^2 dy \\
&\leq \sum_{j=1}^{N}\int_{\mathcal{I}_j^N} K^2\left(\left|\frac{i}{N}-x\right|+\left|\frac{j}{N}-y\right|\right)^2 dy \\
&\leq \sum_{j=1}^{N}\int_{\mathcal{I}_j^N} \frac{4K^2}{N^2} dy \\
&= \frac{4K^2}{N^2}
\end{aligned}$$

furthermore,

$$\int_0^1 (f(x)-f(y))^2 dy \leq \int_0^1 2(f(x)^2+f(y)^2)dy$$

$$= 2f(x)^2 + 2$$

hence inequality (5.4.3) becomes

$$\begin{aligned}\|\mathbb{L}_N f - \mathbb{L} f\|^2 &\leq \frac{8K^2}{N^2} \sum_{i=1}^{N} \int_{B_i^N} \left(f(x)^2 + 1\right) dx \\ &= \frac{16K^2}{N^2}\end{aligned}$$

Consequently,

$$\|\mathbb{L}_N f - \mathbb{L} f\| \leq \frac{4K}{N}.$$

This proves (5.4.1), since this bound holds for all function $f \in L^2([0,1])$ with unit norm. ∎

Proof of inequality (5.4.2). From the triangular inequality and inequality (5.4.1) we get

$$\left|\left|\left|\rho_N^{-1}\widehat{\mathbb{L}}_N - \mathbb{L}\right|\right|\right| \leq \left|\left|\left|\rho_N^{-1}\widehat{\mathbb{L}}_N - \mathbb{L}_N\right|\right|\right| + |||\mathbb{L}_N - \mathbb{L}||| \leq \left|\left|\left|\rho_N^{-1}\widehat{\mathbb{L}}_N - \mathbb{L}_N\right|\right|\right| + \frac{4K}{N}. \tag{5.4.4}$$

Since both $\widehat{\mathbb{L}}_N$ and $\mathbb{L}_N$ have finite rank, it is not hard to see that

$$\begin{aligned}\left|\left|\left|\rho_N^{-1}\widehat{\mathbb{L}}_N - \mathbb{L}_N\right|\right|\right| &= \frac{1}{N}\lambda_{\max}\left(\rho_N^{-1}\widehat{L}^{(N)} - L^{(N)}\right) \\ &= \frac{\rho_N^{-1}}{N}\left\|\widehat{L}^{(N)} - \rho_N L^{(N)}\right\| \\ &\leq \frac{\rho_N^{-1}}{N}\left(\left\|\widehat{D}^{(N)} - \rho_N D^{(N)}\right\| + \left\|\widehat{P}^{(N)} - \rho_N P^{(N)}\right\|\right)\end{aligned} \tag{5.4.5}$$

For some $\varepsilon \in (0, 1/2)$, let N as in Lemma 5.4.2, the maximum expected degree of the random graph represented by $\widehat{P}^{(N)}$ is $\rho_N d_{\text{MAX}}^{(N)} \geq \frac{4}{9}\log(2N/\varepsilon)$ then by Theorem 1 of [5] we get that with probability $1 - \varepsilon$

$$\left\|\widehat{P}^{(N)} - \rho_N P^{(N)}\right\| \leq \sqrt{4\rho_N d_{\text{MAX}}^{(N)} \log(2N/\varepsilon)} \tag{5.4.6}$$

Now we give a similar bound for $\left\|\widehat{D}^{(N)} - \rho_N D^{(N)}\right\|$. Put $\Delta^{(N)} = \widehat{D}^{(N)} - \rho_N D^{(N)}$ and $X^{(N)} = \widehat{P}^{(N)} - \rho_N P^{(N)}$. Then by Bernstein inequality for each $t > 0$

$$\begin{aligned}\mathbb{P}\left(|\Delta_i^{(N)}| > t\right) &= \mathbb{P}\left(\left|\sum_{j=1}^{N} X_{ij}^{(N)}\right| > t\right) \\ &\leq 2\exp\left(\frac{-\frac{1}{2}t^2}{\sum_j \mathbb{V}\text{ar}\left(X_{ij}\right) + \frac{1}{3}t}\right)\end{aligned}$$

$$\leq 2\exp\left(\frac{-\frac{1}{2}t^2}{\rho_N d^{(N)}_{\mathrm{MAX}} + \frac{1}{3}t}\right)$$

where the last inequality follows from the fact that

$$\sum_j \mathbb{V}\mathrm{ar}\left(X_{ij}\right) = \sum_j \rho_N P^{(N)}_{ij}(1-\rho_N P^{(N)}_{ij}) \leq \rho_N \sum_j P^{(N)}_{ij} \leq \rho_N d^{(N)}_{\mathrm{MAX}}.$$

By taking a union bound we get

$$\mathbb{P}\left(\left\|\Delta^{(N)}\right\| > t\right) \leq 2N\exp\left(\frac{-\frac{1}{2}t^2}{\rho_N d^{(N)}_{\mathrm{MAX}} + \frac{1}{3}t}\right).$$

Consider N large enough as in Lemma 5.4.2 then for $t = \sqrt{4\rho_N d^{(N)}_{\mathrm{MAX}} \log\left(2N/\varepsilon\right)}$,

$$\begin{aligned}
\mathbb{P}\left(\left\|\Delta^{(N)}\right\| > \sqrt{4\rho_N d^{(N)}_{\mathrm{MAX}} \log\left(2N/\varepsilon\right)}\right) &\leq 2N\exp\left(-\frac{2\rho_N d^{(N)}_{\mathrm{MAX}} \log(2N/\varepsilon)}{\rho_N d^{(N)}_{\mathrm{MAX}} + \frac{1}{3}\sqrt{4\rho_N d^{(N)}_{\mathrm{MAX}} \log\left(2N/\varepsilon\right)}}\right) \\
&\leq 2N\exp\left(-\frac{2\rho_N d^{(N)}_{\mathrm{MAX}} \log(2N/\varepsilon)}{\rho_N d^{(N)}_{\mathrm{MAX}} + \rho_N d^{(N)}_{\mathrm{MAX}}}\right) \\
&= \varepsilon.
\end{aligned}$$

therefore

$$\left\|\widehat{D}^{(N)} - \rho_N D^{(N)}\right\| = \left\|\Delta^{(N)}\right\| \leq \sqrt{4\rho_N d^{(N)}_{\mathrm{MAX}} \log\left(2N/\varepsilon\right)} \tag{5.4.7}$$

with probability at least $1-\varepsilon$. For $\varepsilon \in (0, 1/2)$ we substitute (5.4.6) and (5.4.7) in (5.4.5) to get

$$\left|\left|\left|\rho_N^{-1}\widehat{\mathbb{L}}_N - \mathbb{L}_N\right|\right|\right| \leq \frac{4\rho_N^{-1}}{N}\sqrt{\rho_N d^{(N)}_{\mathrm{MAX}} \log\left(2N/\varepsilon\right)} \leq 4\sqrt{\frac{\log(2N/\varepsilon)}{\rho_N N}} \tag{5.4.8}$$

with probability at least $1-2\varepsilon$, where we used that $d^{(N)}_{\mathrm{MAX}} \leq N$.

Finally, in inequality (5.4.4) we get that for all $\varepsilon \in (0, 1/2)$ and for N large enough satisfying the conditions of Lemma 5.4.2

$$\left|\left|\left|\rho_N^{-1}\widehat{\mathbb{L}}_N - \mathbb{L}\right|\right|\right| \leq 4\sqrt{\frac{\log(2N/\varepsilon)}{\rho_N N}} + \frac{4K}{N}$$

with probability at least $1-2\varepsilon$. ∎

∎

5.4.2 Convergence of the Fiedler function

A well-known result by Davis and Kahan [6, 29], from spectral theory of operators says that the angle between eigenfunctions of two symmetric matrices is bounded in term of their spectrum:

Theorem 5.4.3 (Davis-Kahan)**.** *Let $\mathbb{L}$ and $\widetilde{\mathbb{L}}$ be two linear operators with eigenvalues $\lambda_1 \leq \lambda_2 \leq \cdots$ and $\tilde{\lambda}_1 \leq \tilde{\lambda}_2 \leq \cdots$ respectively. Consider f and $\tilde{f}$ two eigenfunctions associated the eigenvalues λ_2 and $\tilde{\lambda}_2$ of $\mathbb{L}$ and $\widetilde{\mathbb{L}}$ respectively. If $\|f\| = \|\tilde{f}\| = 1$ and $\langle f, \tilde{f}\rangle \geq 0$. If $\lambda_1 < \lambda_2 < \lambda_3$ (i.e., λ_2 is isolated) then*

$$\left\|\tilde{f} - f\right\| \leq \frac{2\sqrt{2}\left|\left|\left|\widetilde{\mathbb{L}} - \mathbb{L}\right|\right|\right|}{\min\{\lambda_2 - \lambda_1, \lambda_3 - \lambda_2\}}$$

Theorem 3. *Let ϕ and $\hat{\phi}_N$ be the Fiedler functions of $\mathbb{L}$ and $\rho_N\widehat{\mathbb{L}}_N$ respectively such that $\|\phi\| = \|\hat{\phi}_N\| = 1$ and $\langle \phi, \hat{\phi}_N\rangle \geq 0$. For all $\varepsilon \in (0, 1/2)$ and N large enough as in Lemma 5.4.2 we have that with probability $1 - 2\varepsilon$*

$$\left\|\phi - \hat{\phi}_N\right\| \leq 4C\left(\sqrt{\frac{\log(2N/\varepsilon)}{\rho_N N}} + \frac{K}{N}\right) \tag{5.4.9}$$

for some constant $C > 0$.

Proof. The proof follows immediately from Theorem 5.4.3 and Theorem 5.4.1. Note that the Davis-Kahan theorem applies because under our assumptions λ_2 is simple (Theorem 2). ∎

5.5 Bounds on the permutations

For simplicity put $\phi_i = \phi_N(\frac{i}{N})$ and $\hat{\phi}_i = \hat{\phi}_N(\frac{i}{N})$ for all $i = 1, 2, \ldots, N$. It is straightforward to check that the Fiedler vectors of $P^{(N)}$ and $\widehat{P}^{(N)}$ are exactly $(\phi_1, \ldots, \phi_N)^T$ and $\left(\hat{\phi}_1, \ldots, \hat{\phi}_N\right)^T$ respectively (since w_N and $\widehat{w}_N$ are finite rank graphons contructed from $P^{(N)}$ and $\widehat{P}^{(N)}$ respectively). Let $\hat{\pi}^{(N)} \in \mathcal{P}_N$ such that $\hat{\phi}_{\hat{\pi}^{(N)}(1)} \leq \hat{\phi}_{\hat{\pi}^{(N)}(2)} \leq \cdots \leq \hat{\phi}_{\hat{\pi}^{(N)}(N)}$. Since ϕ is monotone and $\hat{\phi}_N$ is close to ϕ (by Theorem 3) we expect the permutation $\hat{\pi}^{(N)}$ to be close to the identity permutation. In what follow, to simplify the notation, we will use the notation σ instead of $\hat{\pi}^{(N)}$. The distance that an element i moved due to a permutation σ (the displacement or drift) is $|\hat{\sigma}(i) - i|$. To count the total displacement, we use the Spearman's footrule distance

$$F_N(\sigma) = \sum_{i=1}^{N} |\sigma(i) - i| .$$

We define the Kendall distance to be

$$D_N(\sigma) = \sum_{i<j} \mathbf{1}_{\{\sigma(i)>\sigma(j)\}}.$$

It was shown in [7] that $D_N(\sigma) \leq F_N(\sigma) \leq 2D_N(\sigma)$. Thus, if N is large enough, these measures are of the same order.

Theorem 4. *Let $\rho_N = \frac{1}{N^\tau}$ for some $\tau \in [0,1)$. For all $\alpha < \frac{1-\tau}{2}$ and $\varepsilon \in (0,1/2)$, if N is large enough then $D_N(\sigma) = \mathcal{O}(N^{2-\frac{2}{3}\alpha})$ with probability at least $1-2\varepsilon$.*

Proof. For all $\beta \in (0,1)$ define

$$E_N := \{(i,j) : i<j \text{ and } \sigma(j)<\sigma(i)\}$$
$$E_N^\beta := \left\{(i,j) : i+N^\beta < j \text{ and } \sigma(j)<\sigma(i)\right\}$$

then $D_N(\sigma) = \mathrm{Card}\,(E_N)$. Besides,

$$\begin{aligned}
2\left\|\hat{\phi}_N - \phi\right\|^2 &= \int_0^1 (\hat{\phi}_N(x)-\phi(x))^2 dx + \int_0^1 (\hat{\phi}_N(y)-\phi(y))^2 dy \\
&= \sum_{i=1}^N \int_{\mathcal{I}_i^N} (\hat{\phi}_i - \phi(x))^2 dx + \sum_{j=1}^N \int_{\mathcal{I}_j^N} (\hat{\phi}_j - \phi(x))^2 dx \\
&= \frac{1}{N} \sum_{1\leq i,j\leq N} \left(\int_{\mathcal{I}_i^N} (\hat{\phi}_i - \phi(x))^2 dx + \int_{\mathcal{I}_i^N} (\hat{\phi}_j - \phi(x+\frac{j-i}{N}))^2 dx \right) \\
&\geq \frac{1}{N} \sum_{(i,j)\in E_N^\beta} \int_{\mathcal{I}_i^N} \left((\hat{\phi}_i - \phi(x))^2 + (\hat{\phi}_j - \phi(x+\frac{j-i}{N}))^2 \right) dx
\end{aligned}$$

for $(i,j) \in E_N^\beta$, we have $j>i$, so $\phi(x+\frac{j-i}{N}) > \phi(x)$ (as ϕ is increasing), and $\hat{\phi}_j < \hat{\phi}_i$ (by definition of E_N^β). Then,

$$(\hat{\phi}_i - \phi(x))^2 + (\hat{\phi}_j - \phi(x+\frac{j-i}{N}))^2 \geq \frac{1}{2}\left(\phi(x+\frac{j-i}{N}) - \phi(x)\right)^2.$$

Therefore,

$$\begin{aligned}
2\left\|\hat{\phi}_N - \phi\right\|^2 &\geq \frac{1}{2N} \sum_{(i,j)\in E_N^\beta} \int_{\mathcal{I}_i^N} \left(\phi(x+\frac{j-i}{N}) - \phi(x)\right)^2 dx \\
&\geq \frac{L^2}{2N^3} \sum_{(i,j)\in E_N^\beta} \int_{\mathcal{I}_i^N} (j-i)^2 dx
\end{aligned}$$

$$\geq \frac{L^2}{2N^{3-2\beta}} \sum_{(i,j)\in E_N^\beta} \int_{\mathcal{I}_i^N} dx$$
$$= \frac{L^2}{2N^{4-2\beta}} \text{Card}\left(E_N^\beta\right)$$

where inequality from line one to two follows from Corollary 5.3.2 and from line two to three follows from $j > i + N^\beta$ (definition of the set E_N^β). Consequently

$$\text{Card}\left(E_N^\beta\right) \leq \frac{4}{L^2} N^{4-2\beta} \left\|\hat{\phi}_N - \phi\right\|^2.$$

Now from Theorem 3 for all $\alpha < \frac{1-\tau}{2}$ we have that $\left\|\hat{\phi}_N - \phi\right\| = \mathcal{O}(N^{-\alpha})$ with probability at least $1 - 2\varepsilon$. Thus

$$\text{Card}\left(E_N^\beta\right) \leq CN^{4-2\beta-2\alpha}.$$

for some constant $C > 0$. On the other hand

$$E \setminus E_N^\beta = \left\{(i,j) : 0 < j - i \leq N^\beta \text{ and } \sigma(j) < \sigma(i)\right\}$$

then $\text{Card}\left(E \setminus E_N\right) \leq N^{\beta+1}$. Choose $\beta = 1 - \frac{2}{3}\alpha$ then

$$D_N(\sigma) = \text{Card}\left(E_N^\beta\right) + \text{Card}\left(E \setminus E_N\right) \leq CN^{4-2\beta-2\alpha} + N^{\beta+1} = (1+C)N^{2-\frac{2}{3}\alpha}$$

as required. ∎

Corollary 5.5.1. *For ρ_N as in Theorem 4 we can conclude that $F_N(\widehat{\pi}^{(N)})/N^2 \to 0$ as $N \to \infty$ with high probability. That is, as expected, $\widehat{\pi}^{(N)}$ is close to the identity permutation for large values of N,*

Bibliography

[1] Jonathan E Atkins, Erik G Boman, and Bruce Hendrickson. A spectral algorithm for seriation and the consecutive ones problem. *SIAM Journal on Computing*, 28(1):297–310, 1998.

[2] Marco Avella-Medina, Francesca Parise, Michael Schaub, and Santiago Segarra. Centrality measures for graphons: Accounting for uncertainty in networks. *IEEE Transactions on Network Science and Engineering*, 2018.

[3] George W Brainerd. The place of chronological ordering in archaeological analysis. *American Antiquity*, 16(4):301–313, 1951.

[4] Huda Chuangpishit, Mahya Ghandehari, Matthew Hurshman, Jeannette Janssen, and Nauzer Kalyaniwalla. Linear embeddings of graphs and graph limits. *Journal of Combinatorial Theory, Series B*, 113:162–184, 2015.

[5] Fan Chung and Mary Radcliffe. On the spectra of general random graphs. *The Electronic Journal of Combinatorics*, 18(1):215, 2011.

[6] Chandler Davis and William Morton Kahan. The rotation of eigenvectors by a perturbation. iii. *SIAM Journal on Numerical Analysis*, 7(1):1–46, 1970.

[7] Persi Diaconis and Ronald L Graham. Spearman's footrule as a measure of disarray. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 262–268, 1977.

[8] Chris Ding. A tutorial on spectral clustering. In *Talk presented at ICML.(Slides available at http://crd. lbl. gov/ cding/Spectral/)*, 2004.

[9] David E Edmunds, AJB Potter, and Charles Alexander Stuart. Non-compact positive operators. *Proc. R. Soc. Lond. A*, 328(1572):67–81, 1972.

[10] Nicolas Flammarion, Cheng Mao, Philippe Rigollet, et al. Optimal rates of statistical seriation. *Bernoulli*, 25(1):623–653, 2019.

[11] Fajwel Fogel, Alexandre d'Aspremont, and Milan Vojnovic. Serialrank: Spectral ranking using seriation. In *Advances in Neural Information Processing Systems*, pages 900–908, 2014.

[12] Fajwel Fogel, Rodolphe Jenatton, Francis Bach, and Alexandre d'Aspremont. Convex relaxations for permutation problems. In *Advances in Neural Information Processing Systems*, pages 1016–1024, 2013.

[13] Delbert Fulkerson and Oliver Gross. Incidence matrices and interval graphs. *Pacific journal of mathematics*, 15(3):835–855, 1965.

[14] Thomas L Gertzen and Martin Grötschel. Flinders petrie, the travelling salesman problem, and the beginning of mathematical modeling in archaeology. *Documenta Mathematica*, pages 199–210, 2012.

[15] K Gustafson and Joachim Weidmann. On the essential spectrum. *Journal of Mathematical Analysis and Applications*, 25(1):121–127, 1969.

[16] Jeannette Janssen and Mahya Ghandehari. An optimization parameter for seriation of noisy data. *arXiv preprint arXiv:1803.10354*, 2018.

[17] David Kendall. Incidence matrices, interval graphs and seriation in archeology. *Pacific Journal of mathematics*, 28(3):565–570, 1969.

[18] David G Kendall. Abundance matrices and seriation in archaeology. *Probability Theory and Related Fields*, 17(2):104–112, 1971.

[19] Gilbert Laporte. The seriation problem and the travelling salesman problem. *Journal of Computational and Applied Mathematics*, 4(4):259–268, 1978.

[20] Monique Laurent and Matteo Seminaroti. The quadratic assignment problem is easy for robinsonian matrices with toeplitz structure. *Operations Research Letters*, 43(1):103–109, 2015.

[21] Innar Liiv. Seriation and matrix reordering methods: An historical overview. *Statistical Analysis and Data Mining: The ASA Data Science Journal*, 3(2):70–91, 2010.

[22] László Lovász. *Large networks and graph limits*, volume 60. American Mathematical Soc., 2012.

[23] WM Flinders Petrie. Sequences in prehistoric remains. *Journal of the Anthropological Institute of Great Britain and Ireland*, pages 295–301, 1899.

[24] Michael Reed. *Methods of modern mathematical physics: Functional analysis*. Elsevier, 2012.

[25] William S Robinson. A method for chronologically ordering archaeological deposits. *American antiquity*, 16(4):293–301, 1951.

[26] Israel Rocha, Jeannette Janssen, and Nauzer Kalyaniwalla. Recovering the structure of random linear graphs. *Linear Algebra and its Applications*, 557:234–264, 2018.

[27] Matteo Seminaroti. Combinatorial algorithms for the seriation problem. Technical report, Tilburg University, School of Economics and Management, 2016.

[28] Ulrike Von Luxburg, Mikhail Belkin, and Olivier Bousquet. Consistency of spectral clustering. *The Annals of Statistics*, pages 555–586, 2008.

[29] Yi Yu, Tengyao Wang, and Richard J Samworth. A useful variant of the davis–kahan theorem for statisticians. *Biometrika*, 102(2):315–323, 2014.