Statistical seriation in non-parametric latent space models: an efficient and optimal algorithm

Abstract

We consider the problem of statistical seriation where one seeks to estimate the ordering between latent positions in $[0,1]$ from pairwise affinities. The observed affinity between a pair of items is modeled as a noisy observation of a function $f(x_i, x_j)$ of the latent positions x_i, x_j of the two items in $[0,1]$. The affinity function f is unknown, and it is only assumed to fulfill some shape constraints ensuring that $f(x,y)$ is large when the distance between x and y is small, and vice-versa. This non-parametric modeling offers a good flexibility to fit data. We shall consider an even more general setting where f fulfills instead a local equivalence between the Euclidean distance in $[0,1]$ and the so-called neighborhood distance. We introduce a computationally efficient procedure that provably recovers the latent ordering of the x_i's with a maximum error of the order of $\sqrt{n\log(n)}$, with high-probability. This rate is proven to be minimax optimal. Our general result can be instantiated to the 1D localization problem [Giraud et al., 2021], leading to new bounds for the maximum error in the localization of the x_i's. This answers an open question raised in [Giraud et al., 2021] about the existence of optimal efficient algorithms.

Contents

1 Introduction

In the seriation problem, one observes similarity measurements between pairs of objects, with a belief that the similarities are characterized by a latent (unknown) ordering between the objects. Specifically, close objects along the ordering share high similarities, while distant objects share relatively low similarities. Seriation then seeks to recover such a latent ordering from the observed similarities.

1.1 Motivation: Original Seriation Problem and its applications

Originally, this problem has its roots in archaeology, in particular for the chronological ordering of graves. Each grave contains artifacts, and the number of common artifacts between

two graves represents their similarity. A high similarity between two graves means that they should be close in the time line. More broadly, such seriation problems arise in many other applications where one wants to order a collection of objects $\{1, \dots, n\}$, when observing a $n \times n$ symmetric matrix $A = [A_{ij}]_{1 \leq i,j \leq n}$, called affinity matrix, which provides similarity measurements between pairs of objects. These similarity measurements A_{ij} can be real valued scores, or they can be binary pieces of information, as when the matrix A encodes a network structure.

Other applications include envelope reduction for sparse matrices [Barnard et al., 1995], reads alignment in *de novo* sequencing [Garriga et al., 2011, Recanati et al., 2017], time synchronization in distributed networks [Elson et al., 2004, Giridhar and Kumar, 2006], interval graph identification [Fulkerson and Gross, 1965], or matchmaking problems [Bradley and Terry, 1952].

The principal assumption on the structure of the affinity matrix A is related to the notion of Robinsonian matrices (also called R-matrices), which are defined as follows. The coefficients of an R-matrix decrease when moving away from the main diagonal, i.e. each row and column of the matrix is unimodal, with a maximum lying on the diagonal of the matrix. In practice, the observed affinity matrix A is a disordered Robinsonian matrix, where the columns and rows have been randomly permuted, and thus do not coincide with a latent ordering. Seriation aims at finding a latent ordering $\pi : [n] \to [n]$ such that $A_\pi = \left[A_{\pi(i)\pi(j)}\right]_{i,j \in [n]}$ is Robinsonian.

It is known that the seriation problem is solved exactly by a spectral algorithm [Atkins et al., 1998], which computes an eigenvector $(v_i)_{i \in [n]}$ of a matrix related to A. The permutation π that sorts the $(v_i)_{i \in [n]}$, i.e. such that $v_{\pi(1)} < \ldots, < v_{\pi(n)}$, turns out to be a latent ordering of A. Note that this result of [Atkins et al., 1998] is reminiscent of popular spectral clusterings [Von Luxburg, 2007], which infer the latent clustering of the data from an eigenvector of a matrix induced by A.

1.2 Problem: Statistical Seriation in General Latent Space Models

Real-world data are often noisy, and the observation A may not be exactly a disordered Robinsonian matrix, though its expectation $\mathbb{E}\, A$ is. In this situation, instead of recovering a latent ordering π perfectly, one aims at building an estimator $\hat{\pi}$ on the noisy data A which only recovers π up to some estimation error. Because the global error of an estimator $\hat{\pi}$, say $\sum_{i=1}^n |\hat{\pi}(i) - \pi(i)|$, provides limited information on each individual error $|\hat{\pi}(i) - \pi(i)|$, it is of practical interest to use instead the maximum error $\max_{i=1,\dots n} |\hat{\pi}(i) - \pi(i)|$.

As probabilistic tools, latent space models [Hoff et al., 2002] are widely-used to study pairwise information data (e.g. networks). In 1D latent space models, the affinity matrix A is assumed to be sampled as follows. The distribution is parametrized by a 1D metric space $(\mathcal{X}, d)$, some (possibly random) latent positions $x_1, \dots, x_n \in \mathcal{X}$ and an affinity function $f : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$. Then, conditionally on $x_1, \dots, x_n$, the upper-diagonal entries A_{ij} of the affinity matrix are sampled independently, with conditional mean $f(x_i, x_j)$. These latent space models encompass many classical models, such as Robinsonian matrices, random geometric graphs, graphon models and monotone Toeplitz matrices (see section 1.4 for these examples).

A simple and general assumption on the affinity $f(x_i, x_j)$ is to decrease as the metric distance $d(x_i, x_j)$ increases. In particular, close points x_i and x_j share a high affinity, whereas distant points share a small affinity. The ordering of the latent points $x_1, \dots, x_n$ then gives a permu-

tation that makes the observed matrix A almost Robinsonian, or more precisely, the expected matrix $\mathbb{E}\,A$ Robinsonian. Indeed, an item i will have a higher affinity with an item j if x_i and x_j are close in the metric space $(\mathcal{X}, d)$. The permutation π induced by the x_i will thus reveal the Robinsonian structure of A. Our aim in this paper is to estimate this permutation π.

However, even with such a shape assumption on f, which ensures that $\mathbb{E}\,A = [f(x_i, x_j)]_{i,j=1,\dots,n}$ is a disordered Robinsonian matrix, there is no guarantee that one can detect the Robinsonian signal in the noisy data A. As an extreme example, one can observe that the constant function $f(x, y) = 1$ fulfills this shape assumption, but leaves no hope of recovering any information on the latent ordering, even in the noiseless case. To circumvent this issue, [Giraud et al., 2021] introduce a bi-Lipschitz assumption, constraining the decay of f with d. On this non-parametric class of bi-Lipschitz functions, it is proved in [Giraud et al., 2021] that the optimal rate of estimation of a latent ordering π is $\max_{i=1,\dots,n} |\hat{\pi}(i) - \pi(i)| = O\left(\sqrt{n \log(n)}\right)$, when the $x_1, \dots, x_n$ uniformly spread in $\mathcal{X}$.

Unfortunately, the algorithm (for bi-Lipschitz functions) in [Giraud et al., 2021] is mainly theoretical, since its time-complexity is super polynomial. This left open the following questions: **Is there any polynomial-time algorithm that achieves the optimal rate over the class of bi-Lipschitz functions ? And when the x_i's do not uniformly spread on $\mathcal{X}$, is it still possible to achieve this rate ?**

A natural idea would be to apply the spectral algorithm of [Atkins et al., 1998]. For instance, [Giraud et al., 2021] show that such a spectral algorithm works under restrictive assumptions about the "uniformity" of the signal $(f(x_i, x_j))_{i,j\in[n]}$, namely that f is geometric (i.e. $f(x_i, x_j)$ only depends on the latent positions x_i, x_j via their distance $d(x_i, x_j)$), and that the x_i's form a uniform sample of $\mathcal{X}$. This paper and others [Rocha et al., 2018, Janssen and Smith, 2020, Cai and Ma, 2022] suggest that spectral algorithms work in such "uniform" restrictive set-ups, but does not in more general settings, such as non-geometric set-up where $f(x_i, x_j)$ varies with the positions x_i, x_j in $\mathcal{X}$ (even if the distance $d(x_i, x_j)$ is constant), or where the $x_1, \dots, x_n$ are not a uniform sample of $\mathcal{X}$.

1.3 Contribution

In the current paper, we recover a latent permutation π induced by the latent positions $x_1, \dots, x_n$, with some high-confidence, simultaneously for all indices $\pi(i)$, $i = 1, \dots, n$. The time complexity of our estimator $\hat{\pi}$ is polynomial, hence the estimator is efficient, which answers positively the first open question (written above in bold). Specifically, we show that, for x_i's uniformly spread in the latent space $\mathcal{X} = [0, 1]$, endowed with the Euclidean distance $d(x, y) = |x - y|$, our estimator $\hat{\pi}$ achieves a maximum error $\max_{i=1,\dots,n} |\hat{\pi}(i) - \pi(i)|$ of the order of $\sqrt{n \log(n)}$, with high-probability. The $\sqrt{n \log(n)}$-rate of estimation is shown to be optimal. In fact, we proves this result under more general assumptions, when f is not necessarily bi-Lipschitz. Indeed, we assume instead a local equivalence between the Euclidean distance in $\mathcal{X} = [0, 1]$ and some distance defined on the columns of the signal matrix $\mathbb{E}\,A = [f(x_i, x_j)]_{i,j=1,\dots,n}$. The latter distance is defined in (3) and is related to the so-called neighborhood distance [Lovász, 2012] which was introduced for the random graphs generated by graphon models [Lovász, 2012].

In order to by-pass the assumption of a uniform spreading of the x_i's and answer the second

open question (written above in bold), we quantify the estimation error of an estimator $\hat{\pi}$ with a slightly more general definition of loss, than the max error $\max_{i=1,\dots n} |\hat{\pi}(i)-\pi(i)|$. Indeed, we say that a permutation $\hat{\pi}$ has an error smaller than ϵ_n, if there exists a latent ordering π such that, for all i, j satisfying $|x_i - x_j| \geq \epsilon_n$, we have the following implication: $\pi(i) < \pi(j) \implies \hat{\pi}(i) < \hat{\pi}(j)$. With this new loss, we reformulate the result stated in the previous paragraph as follows: Our efficient algorithm has the optimal error $\epsilon_n^{opt} := O\left(\sqrt{\log(n)/n}\right)$, when the x_i's spread uniformly on $[0, 1]$. We then show that a slightly refined algorithm achieves the same optimal error ϵ_n^{opt} even when the x_i's do not uniformly spread on $[0, 1]$. To the best of our knowledge, this is the first efficient algorithm that achieves the optimal rate of learning in such a general model, with unknown affinity function f and non-uniform sample $x_1, \dots, x_n$. We hope that these mild assumptions help reduce the gap between practice and theory in the seriation problem.

Our algorithm for ordering the latent points is to proceed to a first partial seriation based on some estimation of the neighborhood distance, and then to expand this partial order in a second stage by comparing each pair of indices at a time. Specifically, in the first step, we estimate the neighborhood distances between latent points using the estimator introduced in [Issartel, 2021]. We then rely on the local equivalence assumption between the neighborhood distance and the Euclidean distance to get a rough ordering of the latent positions. Next, in the second step of the procedure, we compare each pair of items according to their affinities with the initial rough ordering computed in stage 1.

1.4 Related Literature

1.4.1 Examples of Models

Example 1: Random Geometric Graph [Gilbert, 1961, Penrose, 2003, Diaz et al., 2020, De Castro et al., 2017] We observe a random graph with n nodes labelled by $\{1, \dots, n\}$. The graph is encoded into an adjacency matrix $A \in \{0, 1\}^{n\times n}$, by setting $A_{ij} = 1$ if there is an edge between nodes i and j, and $A_{ij} = 0$ otherwise. Let $\mathcal{X}$ be a latent space endowed with a metric d, and let $x_1, \dots, x_n$ be the latent positions of the nodes in $\mathcal{X}$. The edges are sampled independently, with probability $\mathbb{P}[A_{ij} = 1] = g(d(x_i, x_j))$, where $g : [0, \infty) \mapsto [0, 1]$ is a non-increasing function. When $\mathcal{X}$ is one dimensional, this random graph model is an instance of the 1D latent space model where $A_{ij} \in \{0, 1\}$, and $f(x_i, x_j) = g(d(x_i, x_j))$.

Example 2: Graphons and f-Random Graphs [Diaconis and Janson, 2007, Lovász, 2012]. The class of f-random graph models, also called graphon models, encompasses all the distributions on random graphs that are invariant by permutation of nodes. It is parametrized by the set of measurable functions $f : [0, 1]\times[0, 1] \to [0, 1]$. The adjacency matrix A of the graph is sampled as follows. First, n latent positions $x_1, \dots, x_n$ are sampled i.i.d. uniformly on $[0, 1]$. Then, conditionally on $x_1, \dots, x_n$, the edges are sampled independently, with conditional probability $\mathbb{P}[A_{ij} = 1|x_1, \dots, x_n] = f(x_i, x_j)$. The f-random graph model is then an instance of 1D latent space model where $A_{ij} \in \{0, 1\}$, and $\mathcal{X} = [0, 1]$. Unless some additional constraints are imposed on the shape of f, the affinity $f(x_i, x_j)$ may vary arbitrarily with the distance $|x_i - x_j|$.

Example 3: Robinsonian Matrices. A Robinsonian matrix (a.k.a. R-matrix) is any symmetric

matrix $B \in \mathbb{R}^{n\times n}$ whose entries decrease when moving away from the diagonal, i.e. such that $B_{i,j} \geq B_{i+1,j}$ and $B_{i,j} \geq B_{i,j-1}$, for all $1 \leq j \leq i \leq n$. A matrix F is called a disordered R-matrix, when there exists a permutation $\pi : [n] \to [n]$ such that $F_\pi = [F_{\pi(i),\pi(j)}]_{i,j}$ is an R-matrix. The noisy seriation problem [Fogel et al., 2013] amounts to find, from a noisy observation of a disordered R-matrix F, a permutation π such that F_π is an R-matrix. This problem can be recast in the latent space terminology using the latent space $\mathcal{X} = [0, 1]$, with latent positions $x_i = \pi^{-1}(i)/n$ and the affinity function $f(x_i, x_j) = F_{i,j}$. Since F_π is an R-matrix, the function $f(x_{\pi(i)}, x_{\pi(j)}) = f(i/n, j/n)$ is decreasing with the distance $|x_{\pi(i)} - x_{\pi(j)}| = |i - j|/n$.

Example 4: Monotone Toeplitz Matrix: Given one monotone vector $\theta_0 > \theta_1 > \ldots > \theta_{n-1}$, a monotone Toeplitz matrix is defined by $\Theta_{ij} = \theta_{|-(i-1)+(j-1)|}$, $i, j \in [n]$. A matrix F is called a disordered monotone Toeplitz matrix, when there exists a permutation π of $\{1, \ldots, n\}$, such that $F_\pi = [F_{\pi(i),\pi(j)}]_{i,j}$ is a monotone Toeplitz matrix. The noisy seriation problem [Cai and Ma, 2022] amounts to find, from a noisy observation of a disordered monotone Toeplitz matrix F, a permutation π such that F_π is a monotone Toeplitz matrix. Once again, this model can be recast in the latent space terminology using $\mathcal{X} = [0, 1]$, and $x_i = \pi^{-1}(i)/n$ and the affinity function $f(x_i, x_j) = F_{i,j}$.

1.4.2 Related Statistical Seriation

Geometric Setting. The geometric setting is characterized by an affinity function $f(x, y) = g(d(x, y))$,for some real function $g : [0, \infty) \to \mathbb{R}$. Hence, the interaction $f(x_i, x_j)$ only depends on latent positions x_i, x_j via their distance $d(x_i, x_j)$. For statistical seriation in such geometric models, see [Giraud et al., 2021, section 4] and [Natik and Smith, 2021] which study spectral algorithms in line with the original spectral solution of [Atkins et al., 1998]. In fact, the papers [Rocha et al., 2018, Janssen and Smith, 2020, Giraud et al., 2021, Cai and Ma, 2022] on statistical seriation suggest that such spectral methods will not successfully recover the latent ordering in more general models where the setting is not geometric, i.e. when the interaction $f(x_i, x_j)$ depends on the positions x_i, x_j in $\mathcal{X}$. In the current paper, we investigate such general models that are not (necessarily) geometric.

Monotone Toeplitz Matrix. For statistical seriation in disordered monotone Toeplitz matrices, see [Cai and Ma, 2022]. These matrices have been discussed in Example 4 above. A peculiarity of this class of matrices is that the latent order of an object i correlates with the sum of its interactions $(F_{ij})_{j\in[n]}$. Indeed, for an object located at an extremity of a monotone Toeplitz matrix, this sum is relatively low. The efficient algorithm proposed in [Cai and Ma, 2022] leverages this peculiarity to recover the latent ordering, and thus is specific to monotone Toeplitz matrices. Unfortunately, such algorithms will not perform well in more general models, such as the ones we consider in the current paper.

NOTATION: We write $[n]$ the set of integers $\{1, \ldots, n\}$. Given a square matrix F of size $n \times n$ and a permutation $\pi : [n] \to [n]$, we denote the permuted matrix $\left[F_{\pi(i)\pi(j)}\right]_{1\leq i,j\leq n}$ by F_π. We write $|F|_\infty = \max_{i,j\in[n]} |F_{ij}|$ the maximum norm of F. For any finite set R, we denote its cardinal number by $\#R$.

2 Statistical Seriation in Non-Parametric Latent Space Model

2.1 The Model

We observe a realization of a symmetric random matrix $A \in \mathbb{R}^{n\times n}$, whose values on the diagonal are $A_{ii} = 0$. We denote by $F_{ij} = \mathbb{E}[A_{ij}]$ the mean value of A_{ij} and by $E_{ij} = A_{ij} - F_{ij}$ the centered random fluctuation. We assume that A has been generated by a latent space model on $[0,1]$: there exist $x_1, \dots, x_n \in [0,1]$ and a function $f : [0,1] \times [0,1] \to \mathbb{R}$ such that $F_{ij} = f(x_i, x_j)$, so

$$A_{ij} = F_{ij} + E_{ij} = f(x_i, x_j) + E_{ij}, \quad \text{for} \quad 1 \leq i < j \leq n. \tag{1}$$

Both the function f and the latent positions $x_1, \dots, x_n$ are unknown. We emphasize that the latent positions $x_1, \dots, x_n$ are assumed to be fixed[1]. Let us describe our assumptions on the spreading of the latent positions $x_1, \dots, x_n$, the shape of f, and the random fluctuations E_{ij}.

2.1.1 Shape of the Affinity Function

As explained in the introduction, we have in mind that $f(x, y)$ decreases with the Euclidean distance $|x - y|$. Since there is no hope to recover a latent ordering $\pi : [n] \to [n]$ when the function f is flat, a natural idea is to impose a minimal decreasing of $f(x, y)$ with the distance $|x - y|$. In the related problem of latent positions localization [Giraud et al., 2021], the authors also require some Lipschitz continuity of f in their analysis. These two conditions on f are enforced by the Bi-Lipschitz condition described below.

Definition 2.1. Bi-Lipschitz functions. *[Giraud et al., 2021] For any fixed constants $0 < \tilde{\alpha} \leq \tilde{\beta}$, let $\mathcal{BL}[\tilde{\alpha}, \tilde{\beta}]$ be the set of all functions $f : [0,1] \times [0,1] \to \mathbb{R}$ that are symmetric (i.e. $f(x, y) = f(y, x)$ for all $x, y \in [0,1]$) and that satisfy the two following conditions for all $x, y, y' \in [0,1]$,*

$$\begin{aligned} &|f(x,y) - f(x,y')| \leq \tilde{\beta}|y - y'| \ ; \\ &f(x,y') - f(x,y) \geq \tilde{\alpha}\big(|x - y| - |x - y'|\big) \quad \text{if } |x - y| \geq |x - y'| \ . \end{aligned}$$

The first condition enforces Lipschitz continuity and the second one enforces a minimal decreasing of $f(x, y)$ with $|x - y|$. In the geometric case $f(x, y) = g(|x - y|)$ with $g : [0,1] \to \mathbb{R}$ continuously differentiable, these conditions hold when $-\tilde{\beta} \leq g'(t) \leq -\tilde{\alpha}$ for all $t \in [0,1]$.

The non-parametric assumption $f \in \mathcal{BL}[\tilde{\alpha}, \tilde{\beta}]$, with $0 < \tilde{\alpha} \leq \tilde{\beta}$, is simple to interpret and also offers a good flexibility to fit the data. However, it may be naive to assume such a constraint on each pairwise interaction $f(x_i, x_j)$, $i, j \in [n]$, $i \neq j$, so we shall instead consider the following set of three assumptions (2), (4) and (5).

Latent Robinsonian Structure. A symmetric matrix $B := [B_{ij}]_{1\leq,i,j\leq n} \in \mathbb{R}^{n\times n}$ is Robinsonian, if its coefficients decrease when moving away from the diagonal, i.e. if for any $0 \leq i < j \leq n$, one has $B_{jk} - B_{ik} < 0$ for all $k < i$, and $B_{jk} - B_{ik} > 0$ for all $k > j$.

[1] if they were random, our results would apply conditionally on the sampling of $x_1, \dots, x_n$.

As discussed earlier, the seriation problem consists of ordering a disordered Robinsonian matrix (when the rows and columns are not indexed according to the latent ordering). We shall assume accordingly that the signal matrix $F = [F_{ij}]_{1\leq,i,j\leq n} = [f(x_i, x_j)]_{i,j\in[n]}$ is a disordered Robinsonian matrix, where the latent ordering is the one induced by the latent positions $x_1, \ldots, x_n$. This constraint is encapsulated in the following definition of latent Robinsonian structure.

Definition 2.2. Latent Robinsonian Structure. *The symmetric matrix* $[F_{ij}]_{1\leq,i,j\leq n} := [f(x_i, x_j)]_{i,j\in[n]}$ *has a latent Robinsonian structure if, for any* $0 \leq x_i < x_j \leq 1$,

$$F_{jk} - F_{ik} < 0 \text{ for all } x_k < x_i, \quad \text{and} \quad F_{jk} - F_{ik} > 0 \text{ for all } x_k > x_j \ . \tag{2}$$

Hence, for a symmetric matrix F having a latent Robinsonian structure (2), there exists a permutation $\pi : [n] \to [n]$ such that $F_\pi := \left[F_{\pi(i)\pi(j)}\right]_{1\leq i,j\leq n}$ is Robinsonian. This permutation is either the one induced by $x_{\pi(1)} < \ldots < x_{\pi(n)}$ or its reverse $\pi^{rev}(\cdot) = \pi(n - \cdot + 1)$ which satisfies $x_{\pi^{rev}(1)} > \ldots > x_{\pi^{rev}(n)}$.

Note that the latent Robinsonian structure (2) is implied by the bi-Lipschitz definition 2.1, hence it is more general than the bi-Lipschitz condition. Besides, the definition (2) allows F_{ij} to depend on the positions x_i of individuals, hence it encompasses the particular case of the geometric model $F_{ij} = f(|x_i - x_j|)$ that only depends on the distances. In fact, the definition (2) does not even imply that close individuals have a higher affinity than distant people. For example, when $|x_i - x_k| > |x_i - x_j|$, the individual i may still have a higher affinity with k than j, whenever j, k are on different sides of i (e.g. $x_k < x_i < x_k$).

Local equivalence between the neighborhood distance and the Euclidean distance. A key object for analyzing interaction data in latent space models is the neighborhood distance [Lovász, 2012]:

$$d_{nb}(i,j) = \left(\frac{1}{n}\sum_{\ell=1}^{n}(F_{i\ell} - F_{j\ell})^2\right)^{1/2} \quad , \quad \text{for} \quad i,j \in [n] \ , \tag{3}$$

where the quantity $d_{nb}(i,j)$ may be interpreted as measuring the propensity of two individuals i, j to interact with similar individuals. We assume the following "local equivalence" between the neighborhood distance (3) and the Euclidean distance on $[0,1]$: There exists a radius $R > 0$ such that, for all $i, j \in [n]$ satisfying $|x_i - x_j| \wedge d_{nb}(i,j) \leq R$,

$$\alpha|x_i - x_j| \leq d_{nb}(i,j) \leq \beta|x_i - x_j| \ . \tag{4}$$

We emphasize that this equivalence is only assumed at a local level, where the distances $|x_i - x_j| \wedge d_{nb}(i,j)$ are smaller than a radius R. Our analysis will go back and forth between the interactions $[F_{ij}]_{j\in[n]}$ of individuals i and their latent features $x_i \in [0,1]$, via the local relation (4).

Remark that, for a bi-Lipschitz function $f \in \mathcal{BL}[\tilde{\alpha}, \tilde{\beta}]$ and a signal $F_{ij} = f(x_i, x_j)$, one gets $d_{nb}(i,j) \leq \tilde{\beta}|x_i - x_j|$, using the triangle inequality. And similarly, one derives $\tilde{\alpha}'|x_i - x_j| \leq d_{nb}(i,j)$ for some numerical constant $\tilde{\alpha}' \in (0, \tilde{\alpha})$, when the $x_1, \ldots, x_n$ spread uniformly in $[0,1]$. In this sense, the local distance equivalence (4) is implied by the bi-Lipschitz condition.

Strong Robinsonian signal. In the noisy set-up (1), the signal $[F_{ij}]_{1\leq,i,j\leq n}$ may be completely buried under the noise $[E_{ij}]_{1\leq,i,j\leq n}$. Even with a latent Robinsonian structure (2), the matrix F may be too "flat" to expect any recovery of the latent ordering. To exclude this pathological case, it is natural to assume that, for any pair i, j of individuals, say $x_i < x_j$, there is a difference in their affinities with individuals x_k in $[0, x_i]$, who are more similar to x_i than x_j. The same should hold for x_j with individuals x_k in $[x_j, 1]$. Formally, let $\rho, \rho' \in (0, 1/4)$ such that, for all i, j satisfying $|x_i - x_j| \leq \rho$ and $x_i < x_j$, we have

$$\begin{aligned} \sum_{k:\ x_k \leq x_i - \rho'} F_{ik} - F_{jk} &\geq \eta_R |x_i - x_j| n, \quad \text{if } \ x_i \geq 1/2 - \rho/2 \\ \sum_{k:\ x_k \geq x_j + \rho'} F_{jk} - F_{ik} &\geq \eta_R |x_i - x_j| n, \quad \text{if } \ x_j \leq 1/2 + \rho/2. \end{aligned} \tag{5}$$

The condition (5) ensures that, at least in some regions $[0, x_i - \rho']$ or $[x_j + \rho', 1]$ of the space $[0, 1]$, the cumulative interactions of i and j are at a distance greater than $\eta_R |x_i - x_j|$, which is the Euclidean distance between i and j, times a parameter η_R representing the strength of the Robinsonian signal. We emphasize that this assumption is local as it only concerns close points x_i, x_j (at Euclidean distance at most ρ).

Note that, if the affinity function f is a bi-Lipschitz function in $\mathcal{BL}[\tilde{\alpha}, \tilde{\beta}]$, then the matrix of coefficients $F_{ij} = f(x_i, x_j)$ satisfies (5) for some numerical constant $\eta_R \in (0, \tilde{\alpha})$, provided that the x_i spread uniformly on $[0, 1]$.

2.1.2 Spreading of the Latent Positions

When latent positions $x_1, \ldots, x_n$ are gathered in few groups in the latent space $[0, 1]$, the signal matrix $F = [f(x_i, x_j)]_{i,j\in[n]}$ is (almost) piece-wise constant and such data are well-studied in clustering problems. By contrast, the latent positions are well spread over the latent space in the seriation problem. Besides, a common assumption in the literature on latent space models is that the $x_1, \ldots, x_n$ is a uniform sample of $[0, 1]$. Although this uniform spreading assumption makes sense in the seriation problem, it is also restrictive as it does not cover many applications (e.g. in pair-matching problems). Hence, to relax this uniform spreading assumption, we only assume that the $x_1, \ldots, x_n$ cover all segments of length greater than η_s :

$$\sup_{x\in[0,1]} \min_{i\in[n]} |x - x_i| \leq \eta_s \ . \tag{6}$$

The quantity $\eta_s \in [0, 1]$ thus represents some (spatial) sparsity of the sample $x_1, \ldots, x_n$.

2.1.3 Other Assumptions

Upper bound on the signal. Assume that there exists a constant $M > 0$ such that

$$|F|_\infty := \max_{i,j\in[n]} |f(x_i, x_j)| \leq M \ . \tag{7}$$

Sub-Gaussian noise. We assume that the entries E_{ij} for $1 \leq i < j \leq n$ of the noise matrix are independent and follow a sub-Gaussian(1) distribution. It means that, for any

matrix $B \in \mathbb{R}^{n\times n}$ and any real number $t \geq 0$, we have

$$\mathbb{P}\left[\sum_{1\leq i<j\leq n} B_{ij}E_{ij} \geq t\sqrt{\sum_{1\leq i<j\leq n} B_{ij}^2}\right] \leq e^{-t^2/2}. \tag{8}$$

2.2 Objective and Results

For a symmetric matrix F having a latent Robinsonian structure (2), there exists a permutation $\pi : [n] \to [n]$ such that $F_\pi := \left[F_{\pi(i)\pi(j)}\right]_{1\leq i,j\leq n}$ is Robinsonian. This permutation is either the one induced by $x_{\pi(1)} < \ldots < x_{\pi(n)}$ or its reverse $\pi^{rev}(\cdot) = \pi(n - \cdot + 1)$ satisfying $x_{\pi^{rev}(1)} > \ldots > x_{\pi^{rev}(n)}$. We use accordingly the following definition of correct permutation.

Definition 2.3 (Correct Permutation)**.** *Let $x_1, \ldots, x_n$ be n points in $[0, 1]$. We say that a permutation $\pi : [n] \to [n]$ is correct for the latent points $x_1, \ldots, x_n$, when it satisfies either*

$$\forall i, j \in [n] : \qquad \{x_i < x_j\} \longleftrightarrow \{\pi(i) < \pi(j)\} \ ,$$

or

$$\forall i, j \in [n] : \qquad \{x_i > x_j\} \longleftrightarrow \{\pi(i) < \pi(j)\} \ .$$

In other words, π satisfies $x_{\pi^{-1}(1)} < \ldots < x_{\pi^{-1}(n)}$ or the reverse $x_{\pi^{-1}(1)} > \ldots > x_{\pi^{-1}(n)}$.

Our objective in this paper is to estimate a correct permutation from the observed matrix A defined in (1). Because A is a noisy observation of F, there is no hope of recovering exactly a correct permutation, hence this seriation problem requires a definition of estimation error for any estimator $\hat{\pi}$[2]. As discussed in the introduction, we choose a slightly more general definition than the maximum error $\max_{i=1,\ldots n} |\hat{\pi}(i) - \pi(i)|$, in order to handle the situation where $x_1, \ldots, x_n$ is a non-uniform sample of $[0, 1]$.

Definition 2.4 (Estimation Error)**.** *A permutation $\hat{\pi}$ has an (estimation) error smaller than ϵ_n, if there exists a correct permutation π such that, for all i, j satisfying $|x_i - x_j| \geq \epsilon_n$, we have the following implication: $\pi(i) < \pi(j) \implies \hat{\pi}(i) < \hat{\pi}(j)$.*

Remark: in general, the latent positions $x_1, \ldots, x_n$ are not identifiable from the observation A. This comes from the mean $F = \mathbb{E}\, A$ itself which can be induced by many different representations $(x_1, \ldots, x_n, f)$. For a detailed discussion about the representations of F and their non-identifiability, please consult [Giraud et al., 2021, section 2.2]. To be precise about the performance of a procedure $\hat{\pi}$, one should refine Definition 2.4 by considering the class of representations of F, then taking a representation $(x_1, \ldots, x_n, f)$ that minimizes the error ϵ_n of $\hat{\pi}$. Since a formal treatment of representations would be cumbersome and not meaningful for the current problem of estimation, we simply consider a fixed representation $(x_1, \ldots, x_n, f)$ in the sequel. For a special care of the non-identifiability of representations, the interested reader may refer to the work [Giraud et al., 2021] on the estimation of latent positions $x_1, \ldots, x_n$.

[2] An estimator is a function whose input is the observation A, and output is a permutation $\hat{\pi} : [n] \to [n]$. For simplicity, we denote the estimator itself by $\hat{\pi}$.

Although the goal is to estimate a correct permutation, most of our analysis will concern comparison functions $h : [n] \times [n] \to \{-1, 0, 1\}$. In this convenient detour, also used by [Janssen and Smith, 2020], a comparison function h is closely related to a permutation π, if h is a good approximation of the following comparison function:

$$h_\pi(i,j) = \begin{cases} -1 & \text{if } \pi(i) < \pi(j) \\ 1 & \text{if } \pi(i) > \pi(j) \end{cases} . \tag{9}$$

Conversely, a comparison function h naturally leads to the following permutation π_h:

$$\begin{aligned} S_h(i) &= \sum_{j \in [n]} h(i,j) \\ \pi_h(i) &= \#\big\{ j \in [n] : \quad S_h(j) \leq S_h(i) \big\} , \end{aligned} \tag{10}$$

where the ties of S_h are handled arbitrary, so that π_h is a permutation. Observe that $\pi_{h_\pi} = \pi$. In line with Definition 2.4, we introduce the following definitions of correctness and error for a comparison function.

Definition 2.5. *•We say that a comparison function $h : [n] \times [n] \to \{-1, 0, 1\}$ is correct, if there exists a correct permutation $\pi : [n] \to [n]$ such that, for all i, j,*

$$h(i,j) = 1 - 2\mathbb{1}_{\pi(i) < \pi(j)} . \tag{11}$$

•A comparison function $\hat{h}$ has an error smaller than ϵ_n, if it fulfills (11) for all i, j satisfying $|x_i - x_j| \geq \epsilon_n$.

As an intermediate result, Theorem 2.6 deals with the estimation of a correct comparison function. Specifically, the two-step seriation algorithm described in section 3 takes the data A as input, and then outputs a comparison function $\hat{h}$. The accuracy of $\hat{h}$ is guaranteed by the next theorem.

Theorem 2.6. *We assume that the observed matrix A is generated by the noisy model (1) where the signal matrix F has a latent Robinsonian structure (2). We also assume that the five hypotheses (4) to (8) hold, and that the model parameters $\alpha, \beta, \eta_s, R, \rho, \rho'$ and the algorithm parameters $\delta, \delta_1, \delta_2, \delta_3, L$ satisfy some conditions (to be specified later). Then, with probability higher than $1 - 5/n$, the comparison function $\hat{h}$ has an error less than $\frac{10}{\eta_R}\sqrt{\frac{\log(n)}{n}}$.*

Let us give a reformulation of Theorem 2.6. There exists a correct comparison function h, such that, with high probability, the output $\hat{h}$ of the two-step seriation algorithm satisfies

$$\forall i, j \in [n] : \quad |x_i - x_j| \geq \frac{10}{\eta_R}\sqrt{\frac{\log(n)}{n}} \implies \hat{h}(i,j) = h(i,j) .$$

In the next section, we describe the two-step seriation algorithm and give theoretical guarantees for each part of the algorithm. Theorem 2.6 will follow directly from these guarantees.

We use the comparison function $\hat{h}$ and the relation (10) to define the following estimator

$$\hat{\pi} := \pi_{\hat{h}} \tag{12}$$

of a correct permutation. We show in section 3.3 that $\hat{\pi}$ inherits the same accuracy as that of $\hat{h}$, when the spatial sparsity η_s in (6) is small, and more precisely when η_s is smaller than $O\left(\eta_R^{-1}\sqrt{\log(n)/n}\right)$. Unfortunately, there is no guarantee for η_s to be as small as that. Hence, in order to remove this strong dependence on η_s, we add an algorithmic step to $\hat{\pi}$ and get a refined estimator $\hat{\pi}_{ref}$. The next corollary ensures that the accuracy of $\hat{\pi}_{ref}$ is almost the same as that of $\hat{h}$, (almost) independently of the value of η_s.

Corollary 2.7. *Under the assumptions of Theorem 2.6, the permutation $\hat{\pi}_{ref}$ defined in section 3.3 has an error less than $\frac{20}{\eta_R}\sqrt{\frac{\log(n)}{n}}$, with probability at least $1-5/n$.*

Corollary 2.7 follows from Theorem 2.6 and the analysis done in section 3.3.

We show in section 4 that the error $\frac{20}{\eta_R}\sqrt{\frac{\log(n)}{n}}$ is optimal (in the minimax sense).

1D Localization. Let us consider the 1D Localization problem introduced in [Giraud et al., 2021], and more specifically its counterpart on the latent space $\mathcal{X}=[0,1]$ (instead of $\mathcal{X}=$ the unit sphere $\mathcal{C}$ of $\mathbb{R}^2$ as in their paper). The objective is to estimate all the positions $x_1,\dots,x_n$ in maximum error $\max_{i,j\in[n]}|\hat{x}_i-x_i| \wedge \max_{i,j\in[n]}|\hat{x}_i-(1-x_i)|$. Here, the minimum allows to handle the identifiability issue coming from the map $x\mapsto 1-x$ on the latent positions x_i in $[0,1]$.

In this 1D Localization problem, the observed affinity matrix A is assumed to be generated by the noisy model (1), the x_i's to form a uniform sample of $[0,1]$, and the affinity function f to belong to the class $\mathcal{BL}[\tilde{\alpha},\tilde{\beta}]$ of bi-Lipschitz functions (of Definition 2.1), where $0<\tilde{\alpha}\le\tilde{\beta}$ are any numerical constants. It was proved in [Giraud et al., 2021] that the optimal rate of localization in maximum error is $O\left(\sqrt{\log(n)/n}\right)$, when the latent space $\mathcal{X}$ is the unit sphere $\mathcal{C}$ of $\mathbb{R}^2$. Unfortunately, their optimal algorithm has a super-polynomial time-complexity and thus is mainly theoretical. An open question raised in their paper was whether there exists an optimal efficient algorithm for this localization problem. We answer positively this question in the sequel. (Remark: our latent space $\mathcal{X}$ is not the same as in their paper, but we think that the optimal rates are the same in both situations $\mathcal{X}=[0,1]$ and $\mathcal{X}=\mathcal{C}$. Indeed, for the localization problem in $\mathcal{X}=[0,1]$, it should be easy to prove that any estimator has a maximum error of localization greater than $\sqrt{\log(n)/n}$ (up to a multiplicative numerical constant), e.g. by adapting the proof of [Giraud et al., 2021] for $\mathcal{X}=\mathcal{C}$).

We could apply the refined estimator $\hat{\pi}_{ref}$, but since the latent positions uniformly spread in $[0,1]$, it is sufficient to apply the simpler estimator (12) and get an estimation $\hat{\pi}$ of a correct ordering, with an error analyzed in Corollary 3.7. From this estimation, we can derive the estimators $\hat{x}_1,\dots,\hat{x}_n$ by setting $\hat{x}_i=\hat{\pi}(i)/n$ for $i=1,\dots,n$.

Corollary 2.8. *Let $\tilde{\alpha},\tilde{\beta}$ be any numerical constants. Assume that the observed matrix A is generated by the noisy model (1) with sub-Gaussian noise (8), and that the affinity function f belongs to the class $\mathcal{BL}[\tilde{\alpha},\tilde{\beta}]$ of bi-Lipschitz functions, and that the x_i's form a uniform sample of $[0,1]$, and that (7) holds for some numerical constant $M>0$. Then, there exists a constant $C_{\tilde{\alpha}}$, depending only on $\tilde{\alpha}$, such that, with probability at least $1-6/n$, we have*

$$\max_{i,j\in[n]}|\hat{x}_i-x_i| \wedge \max_{i,j\in[n]}|\hat{x}_i-(1-x_i)| \le C_{\tilde{\alpha}}\sqrt{\frac{\log(n)}{n}}\ .$$

Corollary 2.8 follows from an application of Corollary 3.7 which guarantees some performance of $\hat{\pi}$. For details, a sketch of proof of Corollary 2.8 is written in appendix B.

3 Two-Step Seriation Algorithm

The recipe for ordering the latent points is to proceed to a first partial seriation based on the neighborhood distance (3) and then to expand the partial order in a second stage using the strong Robinsonian signal (5).

Stage 1:
(i) *Distance Estimation*: We know from [Issartel, 2021] how to estimate d_{nb} with maximum error $|\hat{d}_{nb}-d_{nb}|_\infty < \delta := O\left(\sqrt{\beta\eta_s} + (\log(n)/n)^{1/4}\right)$ under the assumption (6) on the spreading of latent points, and the assumption (4) on the local distance equivalence. We simply denote the neighborhood distance d_{nb} by d in the rest of the paper.
(ii) *Local Partitioning*: If the neighborhood distance d is locally equivalent to the Euclidean distance as in (4), then for each x_i we will be able to split apart the points at (Euclidean) distance at least $O(\delta)$ of x_i into two groups: those smaller than x_i and those larger than x_i.
(iii) *Consensus*: By symmetry, we are not able to know which one the two groups is on the left of x_i and which one is on the right in $[0,1]$. So we choose arbitrarily a direction, but this choice must be coherent between the n partitions made at each $i \in [n]$. Hence, we apply a consensus step to get a partial order, which compares (without mistakes) all points at (Euclidean) distance larger than $O(\delta)$. In other words, we output a comparison function $\hat{h}$ with an error less than $O(\delta)$.

Stage 2:
In a second stage, we use the latent Robinsonian structure (2) and the strong Robinsonian signal (5) in order to expand the partial ordering to points at (Euclidean) distance as close as $O((\eta_R^{-1}\sqrt{\log(n)/n}) \wedge \delta)$. This expansion is performed by testing sums based on the initial (partial) order of stage 1.

Throughout the section 3, we assume that the observed matrix A is generated by the noisy model $A_{ij} = F_{ij} + E_{ij}$ introduced in (1)*, with $F_{ij} = f(x_i, x_j)$ having a latent Robinsonian structure as in* (2)*. These assumptions are implicit in each proposition of section 3.*

3.1 Stage 1: a First (non-optimal) Seriation based on Neighborhood Distance

Stage 1 is a threefold procedure where we (i) estimate the latent distances, (ii) proceed to a local partitioning for each i, (iii) give a common orientation to these partitions.

3.1.1 Distance Estimation

Recall that the neighborhood distance d_{nb} is simply denoted by d in the sequel. We shall estimate it using the distance estimator $\hat{d}$ introduced in [Issartel, 2021, Eq.(11)]. Hereafter, we call *distance estimation* (DE) algorithm, the data driven procedure in [Issartel, 2021] that

outputs the distance estimates $\left[\hat{d}(i,j)\right]_{i,j\in[n]}$.

Proposition 3.1. *We assume that the signal bound* (7)*, the positions spreading* (6)*, the local distance equivalence (4) and the sub-Gaussian condition* (8) *hold. Also assume that* $\eta_s \leq R$*. Then, with probability at least* $1 - 1/n^3$*, the maximum error of* $\hat{d}$ *is upper bounded by*

$$\max_{i,j\in[n]} \left| d^2(i,j) - \hat{d}^2(i,j) \right| < C \left(M\beta\eta_s + (1 \vee M)\sqrt{\frac{\log(n)}{n}} \right)$$

for some numerical constant $C > 0$*.*

For a proof of Proposition 3.1, see appendix A.1. It follows from Proposition 3.1 that

$$|\hat{d} - d|_\infty < \delta := \sqrt{C} \left(\sqrt{M\beta\eta_s} + (1 \vee \sqrt{M}) \left(\frac{\log(n)}{n} \right)^{1/4} \right) \tag{13}$$

with probability higher than $1 - 1/n^3$.

3.1.2 Local Partitioning (or local comparison)

Local Partitioning (LP)

Input: the matrix $\left[\hat{d}(i,j)\right]_{1\leq i,j\leq n}$ of distance estimates, and some real numbers $\delta_1, \delta_2, \delta_3 > 0$.

1. Build a graph G_i of nodes $\{1, \ldots, n\}$ as follows: link all $k, \ell \in [n] \setminus \{i\}$ fulfilling
$$\hat{d}(k,\ell) \leq \delta_1 \quad \text{and} \quad \hat{d}(i,k) \vee \hat{d}(i,\ell) \geq \delta_2 \ . \tag{14}$$
2. Output all connected components $\mathcal{C}^1(i), \mathcal{C}^2(i), \ldots$ including at least one point x_k such that
$$\hat{d}(x_k, x_i) \geq \delta_3 \ .$$

.

For the sake of brevity, we use the equivalence $k \longleftrightarrow x_k$ between an index k and its associated point x_k, for instance saying that "a point x_k belongs to a connected component of G_i" (though it is, strictly speaking, k that belongs to G_i).

Proposition 3.2. *Assume that the local distance-equivalence* (4) *and the spreading* (6) *of latent positions hold and that*

$$\delta_1 \geq \delta + \beta\eta_s \ , \qquad \delta_2 \geq \delta + \frac{\beta}{\alpha}(\delta_1 + \delta) \ , \qquad R \geq \frac{\eta_s \vee (\delta + \delta_1) \vee (\delta_2 + \delta)}{1 \wedge \alpha} \ .$$

Then, conditionally to the event $|\hat{d} - d|_\infty < \delta$*, the following statements hold for any* $i \in [n]$ *and* $\rho := (\delta_2 + \delta)/\alpha$*:*

1. *all points of a connected component of G_i are on the same side of i;*

2. *all the x_ℓ such that $x_\ell \leq x_i - \rho$ (resp. $x_\ell \geq x_i + \rho$) are in a same connected component linking points all on the same side of x_i.*

In addition if $\eta_s \leq 1/4$ and $\delta + \beta\rho \leq \delta_3 < (R \wedge (\alpha/4)) - \delta$, then

3. *there exist at least one, and at most two, non-empty connected components $\mathcal{C}^1(i)$, $\mathcal{C}^2(i)$, including points x_k such that $\hat{d}(k,i) \geq \delta_3$.*

At this stage, under the above assumptions, there exists for each G_i, at least one, and at most two, non-empty connected components $\mathcal{C}^1(i)$ and $\mathcal{C}^2(i)$ output by the LP algorithm. If $\mathcal{C}^\ell(i)$ for some $\ell \in [2]$ is non-empty, then $\mathcal{C}^\ell(i)$ is on one side of x_i and includes (possibly among others) the points at Euclidean distance at least ρ of x_i. The proof of Proposition 3.2 is in appendix A.2.

3.1.3 Consensus among all Local Comparisons

It is now time to give a common orientation to the n partial orderings induced by the n pairs of components $\mathcal{C}^1(i)$, $\mathcal{C}^2(i)$, $i \in [n]$. Note that $\mathcal{C}^2(i)$ may be the empty set. We use the consensus algorithm (CA) described below. If one of the steps of CA fails, then break (it will not happen under the event $|\hat{d} - d|_\infty < \delta$, which occurs with probability at least $1 - 1/n^3$).

Consensus algorithm (CA)

Input: the connected components $\left[\mathcal{C}^k(i)\right]_{k\in[2],i\in[n]}$ output by the LP algorithm.

1. choose a point i^* having 2 non-empty connected components $\mathcal{C}^1(i^*), \mathcal{C}^2(i^*)$ of size at least $\lfloor \eta_s^{-1}/4 \rfloor$ each. Choose an arbitrary sign: set $C^-(i^*) = \mathcal{C}^1(i^*)$ and $C^+(i^*) = \mathcal{C}^2(i^*)$.

2. for $i = 1, \ldots, n$ and $i \neq i^*$:

   - if G_i has a single connected component $\mathcal{C}^1(i)$, then: if $i \in C^\epsilon(i^*)$ for some $\epsilon \in \{-,+\}$, set $C^{-\epsilon}(i) = \mathcal{C}^1(i)$ and $C^\epsilon(i) = \emptyset$.
   - if G_i has two connected components, then: if $\mathcal{C}^k(i) \cap C^\epsilon(i^*) = \emptyset$ for some $\epsilon \in \{-,+\}$, $k \in [2]$, set $C^{-\epsilon}(i) = \mathcal{C}^k(i)$ and $C^\epsilon(i) = \mathcal{C}^{k'}(i)$ for $k' \neq k$, $k' \in [2]$.

Proposition 3.3. *We assume that the hypotheses of Proposition 3.2 hold and that $\eta_s \leq \rho \leq 1/8$ and $\delta_3 + \delta \leq R \wedge (\alpha/8)$. Conditionally to the event $|\hat{d} - d|_\infty < \delta$, the consensus algorithm runs. With no loss of generality, we can assume that*

$$\operatorname*{arg\,min}_{i \in C^-(i^*) \cup C^+(i^*)} x_i$$

belongs to $C^-(i^*)$, otherwise switch the labels $\epsilon \in \{-,+\}$ of $C^\epsilon(i^*)$. Then, the following inequalities hold for all $i \in [n]$,

$$\forall\, k \in C^-(i), \forall\, k' \in C^+(i): \qquad x_k < x_i < x_{k'} \ .$$

The proof of Proposition 3.3 can be found in appendix A.3.

3.1.4 Comparison Function Estimator

The connected components produced by the CA algorithm are inputs for the CFI algorithm presented below, which will output a comparison function $\hat{h}$ whose accuracy is controlled by Proposition 3.4.

Comparison Function Implementation (CFI)

Input: the connected components $[C^\epsilon(i)]_{\epsilon\in\{-,+\},i\in[n]}$ output by the CA algorithm.

1. for $i, j = 1, \ldots, n$, with $i < j$:
   - if $j \in C^\epsilon(i)$, then set $\hat{h}(i,j) = -\epsilon 1$, otherwise set $\hat{h}(i,j) = 0$.
   - if $\hat{h}(i,j) = 0$ and $i \in C^\epsilon(j)$, then set $\hat{h}(i,j) = \epsilon 1$.
2. for $i, j = 1, \ldots, n$, with $i > j$: set $\hat{h}(i,j) = -\hat{h}(j,i)$.
3. for $i = 1, \ldots, n$: set $\hat{h}(i,i) = 0$.

.

Proposition 3.4. *We assume that the hypotheses of Proposition 3.2 and 3.3. Then, conditionally to the event $|\hat{d}_{nb} - d_{nb}|_\infty < \delta$, the CFI algorithm outputs a comparison function $\hat{h}$ that has an error smaller than ρ.*

The proof of Proposition 3.4 is written in appendix A.4.

3.2 Stage 2: Expansion of the Comparison Function Estimator

In the second stage, we want to compare the pairs i, j that have not been successfully determined in the first stage, that is, those satisfying $\hat{h}(i,j) = 0$ after the call to the CFI algorithm. By Proposition 3.4, such pairs of points are at distance less than $|x_i - x_j| < \rho$. Our objective is to reduce this distance to $(c/\eta_R)\sqrt{\log(n)/n}$. With the assumption (5) in mind, one might think about the natural test below:
If the inequality

$$\max_{\epsilon\in\{-,+\}} \sum_{k\in C^\epsilon(i)} \epsilon(A_{jk} - A_{ik}) \geq L\sqrt{n\log(n)} \tag{15}$$

holds, then it is very likely that $x_j > x_i$, and we set $\hat{h}(i,j) = -1$. Unfortunately, such a test may have cumbersome statistical dependencies with the first stage, as the terms $A_{jk} - A_{ik}$ in

the sum (15) have already been used for the construction of $C^\epsilon(i)$ (over which the sum is). To avoid these complications when comparing two indices i, j, we use the CE algorithm described below, which will go through the first stage again, but with proxy data $A_{-(i,j)}$ which have removed the i^{th}, j^{th} lines of A. Thus, using the square matrix $[A_{st}]_{\substack{s,t\in[n]\\ s,t\neq i,j}}$ instead of A will remove the dependency between stage 1 and 2.

The CE algorithm runs the same (DE and LP) algorithms as before, but without the points x_i, x_j. Accordingly, our analysis will use Propositions 3.1, 3.2 once again, but with respect to $A_{-(i,j)}$. The notations used earlier for A have the following counterparts for $A_{-(i,j)}$: Denote by $d_{-(i,j)}$ the neighborhood distance (3) based on $F_{-(i,j)}$, and by $\hat{d}_{-(i,j)}$ the distance estimator built by the DE algorithm from $A_{-(i,j)}$. For $k = 1, \ldots, n$ and $k \neq i, j$, write $\mathcal{C}^1_{-(i,j)}(k)$, $\mathcal{C}^2_{-(i,j)}(k)$ the connected components output by the LP algorithm, when the using input $\hat{d}_{-(i,j)}$.

Comparaison Expansion (CE)

Input: data matrix A, the connected components $C^\epsilon(i)$, $\epsilon \in \{-,+\}$, $i \in [n]$ output by the CA algorithm, and the estimator $\hat{h}$ computed by the CFI algorithm.

For $i = 1, \dots, n$:

- If $C^+(i) \cup C^-(i) = \emptyset$, then break.
- For $\epsilon = -, +$ such that $C^\epsilon(i) \neq \emptyset$, compute

$$i^\epsilon \in \underset{j \in C^\epsilon(i),\ s.t.\ \hat{d}(i,j) \geq \delta_3}{\arg\min} \hat{d}(i,j) \ .$$

- For $j = 1, \dots, n$ such that $j \neq i$ and $\hat{h}(i,j) = 0$:
  1. Compute the distance estimator $\hat{d}_{-(i,j)}$ using DE algorithm on $A_{-(i,j)}$.
  2. For $\epsilon = -, +$ such that $C^\epsilon(i) \neq \emptyset$:
     (a) Compute $\mathcal{C}^\ell_{-(i,j)}(i^\epsilon)$, $\ell \in [2]$, calling to LP algorithm with input $\hat{d}_{-(i,j)}$.
     (b) Denote by $C^\epsilon_{-(i,j)}(i^\epsilon)$ the component $\mathcal{C}^\ell_{-(i,j)}(i^\epsilon)$, $\ell \in [2]$, contained in $C^\epsilon(i)$.
  3. If

$$\max_{\epsilon \in \{-,+\}} \sum_{k \in C^\epsilon_{-(i,j)}(i^\epsilon)} \epsilon(A_{jk} - A_{ik}) \geq L\sqrt{n \log(n)} \tag{16}$$

     then set $\hat{h}(i,j) = -1$ and $\hat{h}(j,i) = 1$.
  4. Otherwise: if

$$\max_{\epsilon \in \{-,+\}} \sum_{k \in C^\epsilon_{-(i,j)}(i^\epsilon)} \epsilon(A_{ik} - A_{jk}) \geq L\sqrt{n \log(n)} \tag{17}$$

     then set $\hat{h}(i,j) = 1$ and $\hat{h}(j,i) = -1$.

A sum over an empty set is by convention equal to zero. Specifically in point 3 and 4 above, we set $\sum_{k \in C^\epsilon_{-(i,j)}(i^\epsilon)} \epsilon(A_{jk} - A_{ik}) = 0$ when $C^\epsilon_{-(i,j)}(i^\epsilon) = \emptyset$.

Proposition 3.5. *Assume that the Robinsonian properties (2) and (5), the sub-Gaussian noise* (8) *and the hypotheses of Proposition 3.2 and 3.3 hold. Also assume that*

$$\begin{aligned} &\delta_3 \geq \beta\rho + \delta \ , && R \geq (\rho + \eta_s) \vee (\delta_3 + 3\delta + \beta\eta_s) \ , \\ &\alpha > 8(\delta_3 + \delta) \ , && 8\eta_s < 1 \ , \\ &\rho \leq \rho' < \frac{1}{4} \ , && 2(\delta_3 + 3\delta + \beta\eta_s) < \alpha(\rho' - \rho) \ . \end{aligned}$$

Then, conditionally to the event $|d - \hat{d}|_\infty \vee \max_{i,j \in [n], i \neq j} |d_{-(i,j)} - \hat{d}_{-(i,j)}|_\infty \leq \delta$, *the CE*

algorithm runs. Additionally, with probability higher than $1 - 4/n$ *and for* $L > 4$*, it outputs a comparison function* $\hat{h}$ *having an error less than* $\frac{2L}{\eta_R}\sqrt{\frac{\log(n)}{n}}$*.*

Proposition 3.5 is proved in appendix A.5.

The probability of having

$$|d - \hat{d}|_\infty \vee \max_{i,j\in[n],i\neq j} |d_{-(i,j)} - \hat{d}_{-(i,j)}|_\infty > \delta$$

is at most $1/n$ (this follows from a union bound and Proposition 3.1). Hence, the conclusion of Proposition 3.5 holds with probability at least $1 - 5/n$. Theorem 2.6 follows, taking $L = 5$.

3.3 From Comparison Function to Permutation

For any comparison function h, recall that one can define a permutation π_h from h as follows:

$$\begin{aligned} S_h(i) &= \sum_{j\in[n]} h(i,j) \\ \pi_h(i) &= \#\big\{j \in [n] : \quad S_h(j) \leq S_h(i)\big\} \ , \end{aligned} \tag{18}$$

breaking ties arbitrarily (whenever S_h does not take different values), so that π_h is a permutation of $[n]$. The following lemma ensures that π_h is almost as accurate as h, when the sample sparsity η_s is small.

Lemma 3.6. *Assume that the positions spreading* (6) *holds. Then, for any comparison function* h *with an error less than* ν*, the permutation* π_h *in* (18) *has an error less than* $2\nu + \eta_s$*.*

The proof of Lemma 3.6 is written in appendix A.6.

Hence, we can use the output $\hat{h}$ of section 3 and the relation (18) to define the permutation $\hat{\pi} := \pi_{\hat{h}}$. The error of $\hat{h}$ is smaller than $\frac{10}{\eta_R}\sqrt{\frac{\log(n)}{n}}$ by Theorem 2.6, so Lemma 3.6 yields the following corollary.

Corollary 3.7. *Under the assumptions of Theorem 2.6, the permutation* $\hat{\pi} := \pi_{\hat{h}}$ *has an error less than* $\frac{20}{\eta_R}\sqrt{\frac{\log(n)}{n}} + \eta_s$*, with probability at least* $1 - 5/n$*.*

In Corollary 3.7, there are two error terms: the first term η_s represents the spatial sparsity of the latent points $x_1, \ldots, x_n$ in $[0, 1]$. This term can be much greater than the second term $\frac{20}{\eta_R}\sqrt{\frac{\log(n)}{n}}$ which is small when the Robinsonian signal strength η_R is large. It is therefore important to remove the additional error term η_s in Corollary 3.7. We do it using an extra algorithmic step which breaks the ties of $S_{\hat{h}}$ in a non arbitrary fashion.

Permutation Estimator (PE)

Input: comparison function h.

Set $S_h(i) = \sum_{j\in[n]} h(i,j)$ for all $i \in [n]$.

Set $g_h(i) = \#\{j \in [n] : \quad S_h(j) \leq S_h(i)\}$ for all $i \in [n]$. [a]

For $k = 1, \ldots, n$, set $C_k = \#g_h^{-1}(k)$ and define a permutation π_{ref} as follows:

- if $C_k = 1$, set $\pi_{ref}\left(g_h^{-1}(k)\right) = k$.
- if $C_k = 2$, then there exist $s, t \in g_h^{-1}(k)$, $s \neq t$, such that $h(s,t) \in \{0, -1\}$. Set $\pi_{ref}(s) = k-1$ and $\pi_{ref}(t) = k$.
- if $C_k \geq 3$, then
  1. compute $V = \mathrm{DS}(h, g_h^{-1}(k))$ with the notation $V = (v_i)_{i\in[C_k]}$.
  2. set $\pi_{ref}(v_i) = k - (C_k - i)$ for all $i \in [C_k]$.

Output: permutation π_{ref} .

[a] The function g_h is not a permutation when there exist $i \neq j$ such that $S_h(i) = S_h(j)$.

Dichotomic Splitting (DS)

Input: comparison function h, a set E of indices with cardinal number $\#E \geq 2$.

- compute $E_i^-, i, E_i^+ = \mathrm{SIT}(h, E)$.
- if $\#E_i^- \geq 2$, then compute $V^- = \mathrm{DS}(h, E_i^-)$;
  else if $\#E_i^- = 1$, do $V^- = [j]$ for $j \in E_i^-$;
  else $V^- = \emptyset$.
- if $\#E_i^+ \geq 2$, then compute $V^+ = \mathrm{DS}(h, E_i^+)$;
  else if $\#E_i^+ = 1$, do $V^+ = [j]$ for $j \in E_i^+$;
  else $V^+ = \emptyset$.

Output: the list $[V^-, i, V^+]$ if $V^- \neq \emptyset$ and $V^+ \neq \emptyset$,
the list $[i, V^+]$ if $V^- = \emptyset$ and $V^+ \neq \emptyset$,
the list $[V^-, i]$ if $V^- \neq \emptyset$ and $V^+ = \emptyset$.
.

Split In Two (SIT)

Input: comparison function h, set E of indices with cardinal number $\#E \geq 2$.

Take any index $i \in E$ and compute

$$E_i^- = \{j \in E\backslash\{i\} : h(i,j) = 1 \text{ or } h(i,j) = 0\} \ , \quad E_i^+ = \{j \in E : h(i,j) = -1\} \ .$$

Output: E_i^-, i, E_i^+.
.

Lemma 3.8. *For any comparison function h with an error less than ν, the permutation π_{ref} output by the PE algorithm has an error less than 2ν.*

The proof of Lemma 3.8 is written in appendix A.6.

We use the output $\hat{h}$ of section 3 as input in the PE algorithm to get a permutation $\hat{\pi}_{ref}$. Theorem 2.6 and Lemma 3.8 then ensure that that $\hat{\pi}_{ref}$ has an error less than $\frac{20}{\eta_R}\sqrt{\frac{\log(n)}{n}}$. Corollary 2.7 follows.

4 Minimax lower bound

In this section, we prove that the $\eta_R^{-1}\sqrt{\log(n)/n}$ rate in Corollary 2.7 is minimax optimal. Let us consider the observation model $A = F + E$, where we assume that the entries $\{A_{ij} : i < j\}$ follow independent Bernoulli distributions with parameters $F_{ij} = f(x_i, x_j)$. We focus on this particular case of sub-Gaussian distributions for the derivation of a lower bound, as we have in mind random graph applications. The lower bound will hold for this random graph model, and a fortiori for the more general case of sub-Gaussian noise considered in previous sections.

To prove the lower bound, we consider the simpler setting where f_0 is known to the statistician, and is an affine function of the Euclidean distance,

$$f_0(x, y) = \frac{3}{4} - \frac{\tilde{\eta_R}|x - y|}{2}, \quad \text{for all } x, y \in [0, 1],$$

with a parameter $\tilde{\eta_R} \in (0, 1]$. This function f_0 corresponds to a geometric latent model as discussed in the introduction. Let $\boldsymbol{X}_n$ be the set of regular positions $\boldsymbol{X}_n = \big\{\mathbf{x} = (\pi(j)/n)_{1\leq j\leq n} : \pi \in \boldsymbol{\Pi}_n\big\}$, where $\boldsymbol{\Pi}_n$ is the collection of permutations of $[n]$. One readily checks that all assumptions (4) to (8) are satisfied for f_0 and any $\mathbf{x} \in \boldsymbol{\Pi}_n$. Indeed, the condition (6) on the spreading of latent positions holds for any $\eta_s \geq 1/n$. The strong Robinsonian property (5) is satisfied with

$$\eta_R = c\tilde{\eta_R}$$

for some numerical constant $c \in (0, 1)$ and $\rho, \rho' \in (0, 1/8)$. The local bi-Lipschitz assumption (4) is fulfilled with $\alpha = c'\,\tilde{\eta_R}$ and $\beta = c''\,\tilde{\eta_R}$ for some numerical constants $c', c'' \in (0, 1)$ and radius $R \in (0, 1/4)$.

We have a one-to-one correspondence $\boldsymbol{X}_n \cong \boldsymbol{\Pi}_n$, via the application sending any $\mathbf{x} = (x_1, \ldots, x_n) \in \boldsymbol{X}_n$ to $\pi_{\mathbf{x}} \in \boldsymbol{\Pi}_n$ such that $\pi_{\mathbf{x}}(i) = nx_i$ for all $i \in [n]$. This permutation preserves the ordering of the latent positions:

$$\forall i, j \in [n] : \qquad \{x_i < x_j\} \longleftrightarrow \{\pi_{\mathbf{x}}(i) < \pi_{\mathbf{x}}(j)\} \ .$$

Let us define a pseudo-metric on $\boldsymbol{X}_n \cong \boldsymbol{\Pi}_n$ which captures the error ϵ introduced in Definition 2.4 for the estimation of a correct permutation. Given a scalar $\epsilon \in [0, 1]$ and two permutations $\pi_{\mathbf{x}}, \pi_{\mathbf{x}'} \in \boldsymbol{\Pi}_n$, we introduce the event $\mathcal{E}_\epsilon(\pi_{\mathbf{x}}, \pi_{\mathbf{x}'})$ on which $\pi_{\mathbf{x}}, \pi_{\mathbf{x}'}$ agree up to ϵ:

$$\mathcal{E}_\epsilon(\pi_{\mathbf{x}}, \pi_{\mathbf{x}'}) = \Big\{\text{either } \{\pi_{\mathbf{x}}(i) < \pi_{\mathbf{x}}(j)\} \longleftrightarrow \{\pi_{\mathbf{x}'}(i) < \pi_{\mathbf{x}'}(j)\} \text{ for all } i, j \text{ satisfying } |i - j| \geq n\epsilon,$$

$$\text{or } \{\pi_{\mathbf{x}}(i) > \pi_{\mathbf{x}}(j)\} \longleftrightarrow \{\pi_{\mathbf{x}'}(i) < \pi_{\mathbf{x}'}(j)\} \text{ for all } i, j \text{ satisfying } |i - j| \geq n\epsilon\Big\} \ .$$

We take the infimum over ϵ to get the following pseudo-metric

$$D(\pi_{\mathbf{x}}, \pi_{\mathbf{x}'}) = \operatorname*{arg\,min}_{\epsilon \in [0,1]} \left\{ \text{ the event } \mathcal{E}_\epsilon(\pi_{\mathbf{x}}, \pi_{\mathbf{x}'}) \text{ holds} \right\} \ . \tag{19}$$

Recall that $\mathbb{P}_{(\mathbf{x}, f_0)}$ denotes the distribution of A with representation $(\mathbf{x}, f_0)$.

Theorem 4.1. *There exist three positive numerical constants C, C', C'' such that for any $n \geq C'$ and $\eta_R \geq C''\sqrt{\log(n)/n}$, we have the lower bound*

$$\inf_{\hat{\pi}} \sup_{\mathbf{x} \in \mathbf{\Pi}_n} \mathbb{P}_{(\mathbf{x}, f_0)} \left[D(\hat{\pi}, \pi_{\mathbf{x}}) \geq \frac{C}{\eta_R} \sqrt{\frac{\log(n)}{n}} \right] \geq \frac{1}{2} \ ,$$

where the infimum holds over all $\sigma(A)$-measurable functions $\hat{\pi}$.

The proof of Theorem 4.1 is given in appendix C. The lower bound is written over the collection of n-tuples $\mathbf{x} \in \mathbf{\Pi}_n$, which is a subclass of the class considered in our upper bounds (since all $\mathbf{x} \in \mathbf{\Pi}_n$ satisfy the condition (6) for any $\eta_s \geq 1/n$). The lower bound matches the upper bound of Corollary 2.7 up to some multiplicative numerical constant. Therefore, it implies the optimality of the $\eta_R^{-1}\sqrt{\log(n)/n}$ estimation rate of our estimator (in the minimax sense). The fact that the lower bound holds even for a known function (f_0) entails that the $\eta_R^{-1}\sqrt{\log(n)/n}$ estimation rate is not driven by the (absence of) knowledge of the affinity function in our setting (1).

A Proof of Theorem 2.6 (Upper Bound)

A.1 Proof of Proposition 3.1 (Distance Estimation)

For any index $i \in [n]$, we define the index $m_i \in [n]$, $m_i \neq i$, such that x_{m_i} is, with respect to the Euclidean distance, one of the closest points to x_i among the latent points $x_1, \dots, x_n$. Formally, let

$$m_i \in \operatorname*{arg\,min}_{\substack{j \in [n] \\ j \neq i}} \quad |x_i - x_j| \ .$$

A direct adaptation of [Issartel, 2021, Proof of Theorem 7] allows us to derive the following bound on the error of $\hat{d}$, where we only used the sub-Gaussian condition (8) and the signal bound (7). There exists some numerical constant $C > 0$ such that, with probability higher than $1 - 1/n^3$,

$$\max_{i,j \in [n]} \left| d^2(i,j) - \hat{d}^2(i,j) \right| < C \left(M \left[d(i,m_i) + d(j,m_j) \right] + (1 \vee M) \sqrt{\frac{\log(n)}{n}} \right) \ .$$

The spreading (6) of latent positions ensures that $|x_i - x_{m_i}| \leq \eta_s$ for all i. Besides, $\eta_s \leq R$, so that we can use the local distance equivalence (4) to obtain

$$\forall i, j \in [n], \qquad d(i,m_i) + d(j,m_j) \leq \beta \left(|x_i - x_{m_i}| + |x_j - x_{m_j}| \right) \leq 2\beta\eta_s \ ,$$

which concludes the proof of Proposition 3.1. □

A.2 Proof of Proposition 3.2 (Local Partitioning analysis)

Points 1 and 2 of Proposition 3.2 deal with the properties of the connected components, while point 3 is about the existence and the number of connected components. We divide the proof accordingly in two parts.

A.2.1 Points 1 and 2 of Proposition 3.2

For the convenience of the reader, we write below a proposition that encapsulates the two first points of Proposition 3.2.

Proposition. *Assume that the local distance equivalence (4) and the positions spreading* (6) *hold and that*

$$\delta_1 \geq \delta + \beta\eta_s \ , \qquad \delta_2 \geq \delta + \frac{\beta}{\alpha}(\delta_1 + \delta) \ , \qquad R \geq \frac{\eta_s \vee (\delta + \delta_1) \vee (\delta_2 + \delta)}{1 \wedge \alpha} \ .$$

Then, conditionally to the event $|\hat{d} - d|_\infty < \delta$, the following statements hold for $\rho := (\delta_2 + \delta)/\alpha$:

1. *all points of a connected component of G_i are on the same side of i;*

2. *all the x_ℓ such that $x_\ell \leq x_i - \rho$ (resp. $x_\ell \geq x_i + \rho$) are in a same connected component linking points all on the same side of x_i.*

Proof.

Lemma A.1. *If there exists κ, ρ such that the four conditions*

$$d(k,\ell) \leq \delta_1 + \delta \Longrightarrow |x_k - x_\ell| \leq \kappa \tag{20}$$
$$|x_i - x_\ell| \leq \kappa \Longrightarrow d(i,\ell) \leq \delta_2 - \delta \tag{21}$$
$$d(i,\ell) < \delta_2 + \delta \Longrightarrow |x_i - x_\ell| < \rho \tag{22}$$
$$d(\ell,\ell+1) \leq \delta_1 - \delta \tag{23}$$

are met for all i, k, ℓ, then conditionally to the event $|\hat{d} - d|_\infty < \delta$ we have

1. *when (14) is met, then either $x_k, x_\ell > x_i$ or $x_k, x_\ell < x_i$;*
2. *all the x_ℓ such that $x_\ell \leq x_i - \rho$ (resp. $x_\ell \geq x_i + \rho$) are in a same connected component linking points all on the same side of x_i.*

Proof of Lemma A.1: Conditions (20) and (21) together with $|\hat{d} - d|_\infty < \delta$ ensure that if (14) holds, then $|x_k - x_\ell| \leq \kappa$ and $|x_i - x_\ell| \vee |x_i - x_k| > \kappa$, so k and ℓ are on the same side with respect to i.

In addition, condition (22) and $|\hat{d} - d|_\infty < \delta$ yield the implication: if $|x_i - x_\ell| \geq \rho$, then $\hat{d}(i,\ell) \geq \delta_2$. Say for example that $x_\ell \leq x_i - \rho$. As $\hat{d}(\ell,\ell-1) \leq \delta_1$ by (23), we have (14) with $k = \ell - 1$. Hence, x_ℓ and $x_{\ell-1}$ are linked. By induction, we obtain that all points on the left of x_ℓ are linked together. Thus, all points at distance at least ρ from x_i are connected together and they belong to a same side of i. □

Let us now check that the conditions of Lemma A.1 are fulfilled.

Lemma A.2. *Under the assumptions of the proposition above, the conditions from (20) to (23) are fulfilled with*

$$\kappa := \frac{\delta_1 + \delta}{\alpha} \quad \text{and} \quad \rho := \frac{\delta_2 + \delta}{\alpha} \ .$$

Proof of Lemma A.2. First, $d(k,\ell) \leq \delta + \delta_1 \leq R$, so we get from the local distance equivalence (4) that

$$|x_k - x_\ell| \leq \frac{1}{\alpha} d(k,\ell) \leq \frac{\delta + \delta_1}{\alpha} = \kappa \ .$$

For $|x_i - x_\ell| \leq \kappa \leq R$, we have

$$d(i,\ell) \leq \beta\kappa \leq \delta_2 - \delta.$$

For $d(i,\ell) < \delta_2 + \delta \leq R$, we have

$$|x_i - x_\ell| < \frac{\delta_2 + \delta}{\alpha} = \rho.$$

Finally, the positions spreading (6) ensures that $|x_\ell - x_{\ell+1}| \leq \eta_s \leq R$, so we have

$$d(\ell,\ell+1) \leq \beta\eta_s \leq \delta_1 - \delta.$$

The proof of Lemma A.2 is complete, as well as the proof of the proposition above, which corresponds to the points 1 and 2 of Proposition 3.2. □

A.2.2 Point 3 of Proposition 3.2

The proposition below encapsulates the point 3 of Proposition 3.2. It states that for each graph G_i there exists at least one, and at most two, non-empty connected components $\mathcal{C}^1(i), \mathcal{C}^2(i)$ output by the LP algorithm.

Proposition. *Assume that* $\eta_s \leq 1/4$ *and* $\rho \leq R$ *and*

$$\delta + \beta\rho \leq \delta_3 < (R \wedge (\alpha/4)) - \delta.$$

Then, under the assumptions of the proposition of sub-section A.2.1, and for any $i \in [n]$*, there exist at least one, and at most two, non-empty connected components* $\mathcal{C}^1(i)$ *and* $\mathcal{C}^2(i)$ *including points* x_k *such that* $\hat{d}(k,i) \geq \delta_3$*.*

Proof. Given $i \in [n]$, let x_k be a point such that $\hat{d}(k,i) \geq \delta_3$. If $|x_i - x_k| < \rho$, then we have $\hat{d}(k,i) < \beta\rho + \delta$ using the local distance equivalence (4) with $\rho \leq R$. So $\hat{d}(k,i) \geq \delta_3 \geq \delta + \beta\rho$ enforces that $|x_i - x_k| \geq \rho$. We know that, on each side of x_i, all points at distance at least ρ are connected together. Hence, there are at most 2 connected components in G_i containing points x_k that satisfy $\hat{d}(k,i) \geq \delta_3$.

In addition, when $\hat{d}(k,i) < \delta_3 \leq R - \delta$, we have $|x_k - x_i| \leq (\delta_3 + \delta)/\alpha < 1/4$. As either $x_i - x_1$ or $x_n - x_i$ is greater than or equal to $1/4$ for $\eta_s \leq 1/4$, then there exists at least one x_k such that $\hat{d}(i,k) \geq \delta_3$. The proof is complete. □

A.3 Proof of Proposition 3.3 (Consensus Algorithm analysis)

Throughout the section A.3, we assume that $x_1 < x_2 < \ldots < x_n$ (with no loss of generality). This allows us to rewrite the statement of Proposition 3.3 as follows.

Proposition. *We assume that the hypotheses of Proposition 3.2 hold and that* $\eta_s \leq \rho \leq 1/8$ *and* $\delta_3 + \delta \leq R \wedge (\alpha/8)$*. Conditionally to the event* $|\hat{d} - d|_\infty < \delta$*, the consensus algorithm runs. With no loss of generality, we can assume that* $\min C^-(i^*) \cup C^+(i^*) \in C^-(i^*)$*, otherwise switch the labels* ϵ *of* $C^\epsilon(i^*)$*. Then, we have* $\max C^-(i) < i < \min C^+(i)$ *for all* $i \in [n]$*.*

Proof. Let us analyze the points 1 and 2 of the Consensus Algorithm (CA) separately.

• *Point 1 of CA:* The point 1 of the Consensus Algorithm runs, if there exists an index i^* that has two non-empty connected components of cardinal number greater than (or equal to) $\lfloor \eta_s^{-1}/4 \rfloor$.

Let i^* be an index such that x_{i^*} belongs to the interval $[1/2 - \eta_s, 1/2 + \eta_s]$. We know from the spreading assumption (6) that such an index i^* exists. Under the assumption $\eta_s \leq \rho \leq 1/8$, all indices k in

$$V := \{k : x_k \in [0, 1/4]\}$$

satisfy $x_k \leq x_{i^*} - \rho$. The point 2 of Proposition 3.2 then ensures that V is included in a same connected component of G_{i^*}, say $\mathcal{C}$, which is located on one side of i^*. If $\hat{d}(i^*, 1) < \delta_3 \leq R - \delta$, then the local distance equivalence (4) yields $|x_{i^*} - x_1| \leq (\delta_3 + \delta)/\alpha < 1/4$, which contradicts the fact that $|x_{i^*} - x_1| \geq 1/2 - 2\eta_s \geq 1/4$ by the positions spreading (6). Hence, $\hat{d}(i^*, 1) \geq \delta_3$ with $1 \in \mathcal{C}$ (since $1 \in V \subset \mathcal{C}$). It then follows from the point 3 of Proposition 3.2 that $\mathcal{C}$ is

necessarily one of the two connected components $\mathcal{C}^1(i^*)$, $\mathcal{C}^2(i^*)$ released by the LP algorithm. We thus have $V \subset \mathcal{C}^l(i^*)$ for some $l \in [2]$.

Similarly, we can show that the other component $\mathcal{C}^{l'}(i^*)$, $l' \neq l$, $l' \in [2]$, contains $V' := \{k : x_k \in [3/4, 1]\}$. Hence, denoting $\mathcal{C}^1(i^*)$, $\mathcal{C}^2(i^*)$ by $C^+(i^*)$, $C^-(i^*)$, with the convention $\min \mathcal{C}^1(i^*) \cup \mathcal{C}^2(i^*) \in C^-(i^*)$ which means $1 \in C^-(i^*)$, we conclude that

$$V \subset C^-(i^*) \quad \text{and} \quad V' \subset C^+(i^*) \ . \tag{24}$$

The cardinal numbers of V and V' are greater than (or equal to) $\lfloor \eta_s^{-1}/4 \rfloor$ since, by the spreading assumption (6), each interval $[0, 1/4]$, $[3/4, 1]$ contains at least $\lfloor \eta_s^{-1}/4 \rfloor$ different points x_i. Hence, the point 1 of the Consensus Algorithm runs. From the display (24) together with the point 2 of Proposition 3.2, we finally get

$$\max C^-(i^*) < i^* < \min C^+(i^*) \ . \tag{25}$$

- *Point 2 of CA:* Let $i \in \{1, \ldots, n\}$, $i \neq i^*$.

(i) When G_i has a single connected component, we have $\hat{d}(1, i) \wedge \hat{d}(n, i) \leq \delta_3$. This yields $d(1, i) \wedge d(n, i) \leq \delta_3 + \delta \leq R$, and thus

$$\alpha |x_i - x_1| \leq d(1, i) \leq \delta_3 + \delta \ , \quad \text{or} \quad \alpha |x_i - x_n| \leq d(n, i) \leq \delta_3 + \delta \ .$$

We use $\delta_3 + \delta \leq \alpha/8$ and the positions spreading (6) to get

$$x_i \leq x_1 + 1/8 \leq \eta_s + 1/8 \leq 1/4 \ , \quad \text{or} \quad x_i \geq x_n - 1/8 \geq 1 - \eta_s - 1/8 \geq 3/4 \ .$$

Hence, either $i \in V$ or $i \in V'$. It then follows from (24) that $i \in C^\epsilon(i^*)$ for exactly one $\epsilon \in \{\pm\}$. Thus, the Consensus Algorithm outputs $C^{-\epsilon}(i) = \mathcal{C}^1(i)$ and $C^\epsilon(i) = \emptyset$.

(ii) When G_i has two connected components, if for example $i^* < i$, then the connected component $\mathcal{C}^k(i)$ containing x_n has a void intersection with $C^\epsilon(i^*)$ containing x_1. In addition of this empty intersection, exactly one of the three other intersections may be empty too, namely the one between $C^{-\epsilon}(i^*)$ and $\mathcal{C}^{k'}(i)$ for $k' \neq k$. The two remaining intersections are indeed non-empty since $x_1 \in C^\epsilon(i^*) \cap \mathcal{C}^{k'}(i)$ and $x_n \in C^{-\epsilon}(i^*) \cap \mathcal{C}^k(i)$. Regardless of the order in which the algorithm tests (the emptiness of) these four intersections, the output is always $C^{-\epsilon}(i) = \mathcal{C}^k(i)$ and $C^\epsilon(i) = \mathcal{C}^{k'}(i)$.

It follows from the analyses (i) and (ii) above that the point 2 of CA runs. It is then not difficult to check that the output of CA satisfy $\max C^-(i) < i < \min C^+(i)$, using the reference partition (25) and Proposition 3.2.

The proof of Proposition 3.3 is complete. □

A.4 Proof of Proposition 3.4 (CFI analysis)

Throughout the section A.4, we assume that $x_1 < x_2 < \ldots < x_n$, with no loss of generality. Let h be the correct comparison function defined by $h(i, j) = 1 - 2\mathbb{1}_{x_i < x_j}$ for all i, j.

The correct comparison matrix $[h(i, j)]_{i,j \in [n]}$ is anti-symmetric by definition, hence (in the point 2 of CFI algorithm) we set the lower triangle of the matrix $\left[\hat{h}(i, j)\right]_{i,j \in [n]}$ as equal

to the opposite of its upper triangle. We analyze this upper triangle in the following. Let $i, j \in [n]$, $i < j$, such that $|x_i - x_j| \geq \rho$.

• 1st *Case: G_i has two non-empty connected components $C^-(i)$ and $C^+(i)$.* We know from Proposition 3.2 that j belongs to one of these two components (since $|x_i - x_j| \geq \rho$). The output of the CFI algorithm is therefore $\hat{h}(i,j) = -\epsilon 1$. Proposition 3.3 then ensures that $\hat{h}(i,j)$ coincides with the value $h(i,j) = 1 - 2\mathbb{1}_{i<j}$ of the correct comparison function h.

• 2nd *Case: G_i has exactly one non-empty connected component $C^\epsilon(i)$.* If $j \in C^\epsilon(i)$, then the CFI algorithm outputs $\hat{h}(i,j) = -\epsilon 1$ again, and Proposition 3.3 allows us to conclude as in the 1st case.

Assume therefore that $j \notin C^\epsilon(i)$. It is sufficient to analyze the case $\epsilon = -$ (the case $\epsilon = +$ being symmetric). Thus, $C^-(i)$ is the unique non-empty connected component of G_i, and $j \notin C^-(i)$. If $x_j < x_i$, then $x_j - \rho \leq x_i$ (since $|x_i - x_j| \geq \rho$), so the point 2 of Proposition 3.2 yields $j \in C^-(i)$, which is a contradiction. Hence, $x_j > x_i$.

If $x_j < 1/2$, then $x_i < 1/2$. If $\hat{d}(i,n) < \delta_3 \leq R - \delta$, then the local distance equivalence (4) yields $|x_i - x_n| \leq (\delta_3 + \delta)/\alpha < 1/4$, which contradicts the fact that $|x_i - x_n| \geq (1 - \eta_s) - 1/2 \geq 1/4$, deduced from the positions spreading (6). Hence, $\hat{d}(i,n) \geq \delta_3$, leading to $n \in C^+(i)$, which is a contradiction with $C^+(i) = \emptyset$. Therefore, $x_j \geq 1/2$.

If $\hat{d}(j,1) < \delta_3 \leq R - \delta$, then (4) yields $|x_j - x_1| \leq (\delta_3 + \delta)/\alpha < 1/4$, which contradicts the fact that $|x_j - x_1| \geq 1/2 - \eta_s \geq 1/4$. Hence, $\hat{d}(j,1) \geq \delta_3$, implying that the component $C^-(j)$ is non-empty with $1 \in C^-(j)$. Since $x_j - \rho \geq x_i$, the point 2 of Proposition 3.2 ensures that $i \in C^-(j)$. The CFI algorithm therefore outputs $\hat{h}(i,j) = \epsilon 1 = -1$, which matches the correct value $h(i,j) = 1 - 2\mathbb{1}_{i<j}$. This concludes the analysis of the 2nd case.

The proof of Proposition 3.4 is complete. □

A.5 Proof of Proposition 3.5 (CE analysis)

Throughout the section A.5, we assume that $x_1 < x_2 < \ldots < x_n$ (with no loss of generality), and we denote by h the following correct comparison function $h(i,j) = 1 - 2\mathbb{1}_{x_i < x_j}$ for all i, j.

We start with three lemmas, the first one controlling the random fluctuations of the noise.

Lemma A.3. *Assume that the sub-Gaussian error* (8) *holds. Then, with probability higher than* $1 - 4/n$,

$$\max_{\epsilon,i,j,\ell} \frac{1}{\sqrt{2|C^\epsilon_{-(i,j)}(\ell)|}} \Big| \sum_{k \in C^\epsilon_{-(i,j)}(\ell)} (E_{ik} - E_{jk}) \Big| < 2\sqrt{2\log(n)} \tag{26}$$

Proof Lemma A.3. The sets $C^-_{-(i,j)}(\ell)$, $C^+_{-(i,j)}(\ell)$ are independent of $E_{ik} - E_{jk}$, $k \in [n]$. Then, for any $\epsilon \in \{-,+\}$ we have that, conditionally on $C^\epsilon_{-(i,j)}(\ell)$,

$$\mathbb{P}\left[\frac{1}{\sqrt{2|C^\epsilon_{-(i,j)}(\ell)|}} \left| \sum_{k \in C^\epsilon_{-(i,j)}(\ell)} (E_{ik} - E_{jk}) \right| \geq t \right] \leq 2e^{-t^2/2} \ ,$$

using the sub-Gaussian property (8). The lemma follows from the choice of value $t = 2\sqrt{2\log(n)}$, together with a union bound over all possible indices i, j, l in $[n]$ and $\epsilon \in \{-, +\}$. □

The next lemma shows that x_i is close to x_{i^+}, x_{i^-}.

Lemma A.4. *Assume that the spreading* (6) *and the local distance equivalence* (4) *hold, and that* $\delta_3 \geq \beta\rho + \delta$ *and* $R \geq (\rho + \eta_s) \vee (\delta_3 + 3\delta + \beta\eta_s)$. *Also assume that the hypotheses of Proposition 3.2 and Proposition 3.3 hold. Conditionally to the event* $|\hat{d} - d|_\infty < \delta$, *the following holds for any* $i \in [n]$. *If* i^ϵ *exists for* $\epsilon \in \{-, +\}$, *then*

$$\alpha|x_i - x_{i^\epsilon}| \leq \delta_3 + \beta\eta_s + 3\delta \ . \tag{27}$$

Proof of Lemma A.4. We only prove the bound for x_{i^+}, the one for x_{i^-} being similar. Set

$$b^+ = \min\left\{ j \in C^+(i) : \hat{d}(i,j) \geq \delta_3 \right\}.$$

Assume that $i^+ \in [n]$ exists. Then, $C^+(i)$ is non-empty, and by definition of $C^+(i)$ there exists k such that $\hat{d}(i,k) \geq \delta_3$. Hence, $b^+ \in [n]$ exists. We also know from Proposition 3.2 that the set $C^+(i)$ includes all points x_k satisfying $x_k - \rho \geq x_i$.

- If $b^+ - 1 \notin C^+(i)$, then $x_{b^+-1} - x_i < \rho$. Proposition 3.3 ensures that $x_{b^+} > x_i$, so we have $0 < x_{b^+} - x_i = (x_{b^+} - x_{b^+-1}) + (x_{b^+-1} - x_i) < \eta_s + \rho$, using the spreading assumption (6). As $\rho + \eta_s \leq R$, the local distance equivalence (4) yields

  $$\hat{d}(i^+, i) \leq \hat{d}(b^+, i) \leq \beta(\rho + \eta_s) + \delta \leq \delta_3 + \beta\eta_s \ ,$$

  where we used the definition of i^+ in the first inequality, the event $|\hat{d} - d|_\infty < \delta$ in the second one, and $\delta_3 \geq \beta\rho + \delta$ in the last one.

- If $b^+ - 1 \in C^+(i)$, then by definition of b^+ we have $\hat{d}(b^+ - 1, i) < \delta_3$. It then follows from the triangular inequality that $d(b^+, i) \leq d(b^+, b^+ - 1) + d(b^+ - 1, i) \leq \beta\eta_s + \delta_3 + \delta$, using the spreading assumption (6) and the local distance equivalence (4). Hence, we get from the definition of i^+ that

  $$\hat{d}(i^+, i) \leq \hat{d}(b^+, i) \leq \beta\eta_s + \delta_3 + 2\delta \ .$$

To conclude, we have

$$d(i^+, i) \leq \delta_3 + \beta\eta_s + 3\delta \leq R \ ,$$

hence the bound (27) follows from the local distance equivalence (4). □

In the following lemma, we give some conditions on the positions x_i, x_j so that the indices i^ϵ exist and the sets $C^\epsilon(i)$, $C^\epsilon_{-(i,j)}(i^\epsilon)$ are non-empty.

Lemma A.5. *Assume that the hypotheses of Proposition 3.2, 3.3 and Lemma A.4 hold. Also assume that* $\delta_3 \geq \beta\rho+\delta$ *and* $R \geq (\delta_3+\delta)\vee\rho$ *and* $\delta_3+\delta < \alpha/8$ *and* $\eta_s < 1/8$ *and* $\rho \leq \rho' < 1/4$ *and* $\delta_3+3\delta+\beta\eta_s < \alpha(\rho'-\rho)/2$. *Conditionally to the event* $|\hat{d}-d|_\infty < \delta$, *we have the following implication for any* $i,j \in [n]$ *such that* $|x_i - x_j| < \rho$.
If $x_i \leq 1/2 + \rho/2$, *then*

$$(a)\ \ C^+(i) \neq \emptyset\ , \qquad\qquad (b)\ \ i^+ \textit{exists}\ ,$$

and conditionally to the event $|\hat{d}_{-(i,j)} - d_{-(i,j)}|_\infty < \delta$,

$$(c)\ \ C^+_{-(i,j)}(i^+) \neq \emptyset\ , \qquad\qquad (d)\ \ \big\{k:\ \ x_k \geq x_i \vee x_j + \rho'\big\} \subset C^+_{-(i,j)}(i^+)\ .$$

Similarly: If $x_i \geq 1/2 - \rho/2$, *then*

$$\text{(a')}\ \ C^-(i) \neq \emptyset\ , \qquad\qquad \text{(b')}\ \ i^- \textit{exists}\ ,$$
$$\text{(c')}\ \ C^-_{-(i,j)}(i^-) \neq \emptyset\ , \qquad \text{(d')}\ \ \big\{k:\ \ x_k \leq x_i \wedge x_j - \rho'\big\} \subset C^-_{-(i,j)}(i^-)\ .$$

Proof of Lemma A.5. Given $i \in [n]$, assume that $x_i \leq 1/2 + \rho/2$ and $|\hat{d} - d|_\infty < \delta$.

(a) If $\hat{d}(i,n) < \delta_3 \leq R-\delta$, then the local distance equivalence (4) yields $|x_i - x_n| \leq (\delta_3+\delta)/\alpha < 1/4$, which contradicts the fact that $|x_i - x_n| \geq (1-\eta_s) - (1/2+\rho/2) \geq 1/4$. Hence, $\hat{d}(i,n) \geq \delta_3$, leading to $n \in C^+(i)$ and $C^+(i) \neq \emptyset$.

(b) When $C^\epsilon(i) \neq \emptyset$, the index i^ϵ exists by construction of the set $C^\epsilon(i)$. Hence, the existence of i^+ follows directly from the point (a) of Lemma A.5.

Given $j \in [n] \setminus \{i\}$ such that $|x_i - x_j| < \rho$, assume (in the sequel) that $|\hat{d}_{-(i,j)} - d_{-(i,j)}|_\infty < \delta$.

(c) By Proposition 3.3, we have $x_k > x_i$ for all $k \in C^+(i)$, in particular $x_{i^+} > x_i$ as $i^+ \in C^+(i)$. It follows from this and the equation (27) of Lemma A.4 that

$$0 < x_{i^+} - x_i \leq \frac{\delta_3 + 3\delta + \beta\eta_s}{\alpha} < \frac{\rho' - \rho}{2} < \frac{1}{8}\ . \tag{28}$$

If $\hat{d}_{-(i,j)}(i^+, n) < \delta_3 \leq R - \delta$, then the local distance equivalence (4) yields $|x_{i^+} - x_n| \leq (\delta_3 + \delta)/\alpha < 1/8$, which contradicts the fact that $|x_{i^+} - x_n| \geq |x_n - x_i| - |x_{i^+} - x_i| \geq |(1-\eta_s) - (1/2+\rho/2)| - 1/8 \geq 1/8$. Hence, $\hat{d}_{-(i,j)}(i^+, n) \geq \delta_3$, thus leading to $n \in C^+_{-(i,j)}(i^+)$ and $C^+_{-(i,j)}(i^+) \neq \emptyset$.

Above, the quantity $\hat{d}_{-(i,j)}(i^+, n)$ and the set $C^+_{-(i,j)}(i^+)$ are well-defined since $\{i,j\} \cap \{i^+, n\} = \emptyset$. Indeed, we have $i \neq n$ and $j \neq n$ because $x_i \leq 1/2 + \rho/2 < 5/8$ and $x_j \leq x_i + \rho < 7/8$ while $x_n \geq 1 - \eta_s > 7/8$. We also have

$$i \neq i^+ \qquad \textit{and} \qquad j \neq i^+\ , \tag{29}$$

since i^+ satisfies $\hat{d}(i, i^+) \geq \delta_3 \geq \beta\rho + \delta$ which implies $|x_i - x_{i^+}| \geq \rho$ (indeed, we know from the local distance equivalence (4) that the contraposition holds: $|x_i - x_{i^+}| < \rho \leq R \implies \hat{d}(i, i^+) < \beta\rho + \delta$). The second relation $j \neq i^+$ then follows from the fact that $|x_i - x_j| < \rho$.

(d) Given an index $k \in [n]$ satisfying $x_k \geq x_i + \rho'$, we get from (28) that

$$x_k - x_{i^+} = (x_k - x_i) - (x_{i^+} - x_i) > (\rho' + \rho)/2 \geq \rho\ ,$$

which implies $k \in C^+_{-(i,j)}(i^+)$ by Proposition 3.2 and 3.3, whenever $C^+_{-(i,j)}(i^+) \neq \emptyset$. As $C^+_{-(i,j)}(i^+) \neq \emptyset$ is guaranteed by the point (c) of Lemma A.5, we obtain $\{k : \ x_k \geq x_i + \rho'\} \subset C^+_{-(i,j)}(i^+)$, and a fortiori $\{k : \ x_k \geq x_i \vee x_j + \rho'\} \subset C^+_{-(i,j)}(i^+)$.

This concludes the proof of the points (a) to (d) for $x_i \leq 1/2 + \rho/2$. By symmetry, the points (a') to (d') for $x_i \geq 1/2 - \rho/2$ follow from a similar proof. □

We are now ready to prove Proposition 3.5, which is rewritten below for the convenience of the reader.

Proposition. *Assume that the Robinsonian properties (2) and (5), the sub-Gaussian noise* (8) *and the hypotheses of Proposition 3.2 and 3.3 hold. Also assume that the hypotheses of Lemma A.4 and A.5 hold. Then, conditionally to the event* $|d - \hat{d}|_\infty \vee \max_{i,j\in[n],i\neq j} |d_{-(i,j)} - \hat{d}_{-(i,j)}|_\infty \leq \delta$*, the CE algorithm runs. Additionally, with probability higher than* $1 - 4/n$ *and for* $L > 4$*, it outputs a comparison function* $\hat{h}$ *having an error less than* $\frac{2L}{\eta_R}\sqrt{\frac{\log(n)}{n}}$.

Proof. Given $i \in [n]$, le us show that (a) the CE algorithm runs until the point 3 of the pseudo-code (at least), and (b) the output $\hat{h}$ of CE algorithm has a small error.

(a) *The CE algorithm runs until point 3:*

Assume that the event $|d - \hat{d}|_\infty \leq \delta$ holds. If $x_i \leq 1/2 + \rho/2$, then Lemma A.5 ensures that $C^+(i) \neq \emptyset$ and i^+ exists. Otherwise, when $x_i > 1/2 + \rho/2$, we obviously have $x_i \geq 1/2 - \rho/2$, and Lemma A.5 yields $C^-(i) \neq \emptyset$ and the existence of i^-. In any case, $C^-(i) \cup C^+(i) \neq \emptyset$, so the CE algorithm does not break at the first bullet point of the pseudo-code, and will run until the third bullet point, where it will scan the indices j such that $\hat{h}(i,j) = 0$. Proposition 3.4 implies that such j satisfy

$$|x_i - x_j| < \rho \ . \tag{30}$$

At the point 1 of the third bullet point, the distance estimator $\hat{d}_{-(i,j)}$ is computed without the points x_i, x_j. Conditionally to the event $|d - \hat{d}|_\infty \vee \max_{i,j\in[n],i\neq j} |d_{-(i,j)} - \hat{d}_{-(i,j)}|_\infty \leq \delta$, we analyze the rest of the algorithm.

At the point 2, the algorithm takes an ϵ such that $C^\epsilon(i) \neq \emptyset$, which exists (as proved in the before last paragraph). With no loss of generality, we assume that $C^+(i) \neq \emptyset$. The algorithm then computes the components $\mathcal{C}^\ell_{-(i,j)}(i^+)$, $\ell \in [2]$, which enjoy the theoretical properties stated in Proposition 3.2. These components are well-defined, in particular i^+ is not equal to i or j, as checked in (29). By definition of i^+, we know that $\hat{d}(i, i^+) \geq \delta_3$, which implies $|x_i - x_{i^+}| \geq \rho$. Hence, $x_k - \rho \geq x_i$ for all $k \in \{i^+, \dots, n\}$, so Proposition 3.2 and 3.3 yield $\{i^+, \dots, n\} \subset C^+(i)$. As $\{i^+, \dots, n\}$ contains one of the two components $\mathcal{C}^1_{-(i,j)}(i^+)$, $\mathcal{C}^2_{-(i,j)}(i^+)$, we finally get for some $\ell \in [2]$ that

$$C^+_{-(i,j)}(i^+) := \mathcal{C}^\ell_{-(i,j)}(i^+) \subset \{i^+, \dots, n\} \subset C^+(i) \ , \tag{31}$$

where we denoted $\mathcal{C}^\ell_{-(i,j)}(i^+)$ by $C^+_{-(i,j)}(i^+)$.

The same applies to the case $\epsilon = -$ when $C^-(i) \neq \emptyset$, meaning that we have, for some $\ell' \in [2]$,

$$C^-_{-(i,j)}(i^-) := \mathcal{C}^{\ell'}_{-(i,j)}(i^-) \subset \{0, \dots, i^-\} \subset C^-(i) \ . \tag{32}$$

Hence, we proved that the CE algorithm runs until the point 3 of the pseudo-code (at least), where it will (possibly) update the estimator $\hat{h}$ through the tests (16-17). These tests are trivial whenever the sums involved in (16-17) are by convention equal to zero, due to being over empty sets $C^-_{-(i,j)}(i^-), C^+_{-(i,j)}(i^+)$. Fortunately, this does not happen under our assumptions, since Lemma A.5 guarantees $C^-_{-(i,j)}(i^-) \cup C^+_{-(i,j)}(i^+) \neq \emptyset$.

(b) *The Output $\hat{h}$ has a Small Error:*

It suffices to show that, for all j satisfying (30), the value $\hat{h}(i,j)$ matches the correct value $h(i,j) = 1 - 2\mathbb{1}_{x_i < x_j}$.

• 1^{st} *Case* $x_i < x_j$. Observe that the CE algorithm outputs the correct value $h(i,j)$ if the inequality (16) is satisfied. Hence, all we need is to show that this inequality holds w.h.p.. Under the event (26) which holds with probability higher than $1 - 4/n$ (by Lemma A.3), we have the following lower bound for all $\epsilon \in \{-, +\}$,

$$\begin{aligned}
\sum_{k \in C^\epsilon_{-(i,j)}(i^\epsilon)} \epsilon(A_{jk} - A_{ik}) &= \sum_{k \in C^\epsilon_{-(i,j)}(i^\epsilon)} \epsilon(F_{jk} - F_{ik}) + \sum_{k \in C^\epsilon_{-(i,j)}(i^\epsilon)} \epsilon(E_{jk} - E_{ik}) \\
&\geq \sum_{k \in C^\epsilon_{-(i,j)}(i^\epsilon)} \epsilon(F_{jk} - F_{ik}) - \max_{\epsilon \in \{-,+\}} \Big| \sum_{k \in C^\epsilon_{-(i,j)}(i^\epsilon)} (E_{jk} - E_{ik}) \Big| \\
&\geq \sum_{k \in C^\epsilon_{-(i,j)}(i^\epsilon)} \epsilon(F_{jk} - F_{ik}) - 4\sqrt{n \log(n)} \ .
\end{aligned} \tag{33}$$

To lower bound the sum in (33), we are going to use the Robinson property (5) for all j satisfying (30) :

$$\begin{aligned}
\sum_{k: \ x_k \leq x_i - \rho'} F_{ik} - F_{jk} \geq \eta_R |x_i - x_j| n, \quad \text{if} \ \ x_i \geq 1/2 - \rho/2 \\
\sum_{k: \ x_k \geq x_j + \rho'} F_{jk} - F_{ik} \geq \eta_R |x_i - x_j| n, \quad \text{if} \ \ x_j \leq 1/2 + \rho/2 \ .
\end{aligned}$$

If $x_i \vee x_j \leq 1/2 + \rho/2$, then Lemma A.5 ensures that $\{k : \ x_k \geq x_j + \rho'\} \subset C^+_{-(i,j)}(i^+)$, thus yielding

$$\begin{aligned}
\max_{\epsilon \in \{-,+\}} \sum_{k \in C^\epsilon_{-(i,j)}(i^\epsilon)} \epsilon(F_{jk} - F_{ik}) \geq \sum_{k \in C^+_{-(i,j)}(i^+)} (F_{jk} - F_{ik}) \geq \sum_{k: \ x_k \geq x_j + \rho'} (F_{jk} - F_{ik}) \\
\geq \eta_R |x_i - x_j| n \ ,
\end{aligned}$$

where in the before last inequality we used the latent Robinsonian structure (2), i.e. $F_{jk} - F_{ik} > 0$ for all $x_k > x_j$, and the fact that all $k \in C^+_{-(i,j)}(i^+)$ satisfy $x_k > x_j$. (For a check of this fact, see the 2^{nd} case below).

From this last display and (33) we get that the following inequality

$$\max_{\epsilon\in\{-,+\}} \sum_{k\in C^{\epsilon}_{-(i,j)}(i^{\epsilon})} \epsilon(A_{jk} - A_{ik}) \geq \eta_R |x_i - x_j| n - 4\sqrt{n\log(n)} \tag{34}$$

holds with probability higher than $1 - 4/n$.

If $x_i \vee x_j > 1/2 + \rho/2$, then we have $x_i \geq 1/2 - \rho/2$ using (30). Lemma A.5 then ensures that $\{k : \ x_k \leq x_i - \rho'\} \subset C^{-}_{-(i,j)}(i^{-})$, leading to (34) once again.

Hence, (34) holds (whatever the value of $x_i \vee x_j$). The right-hand side of (34) is greater than $L\sqrt{n\log(n)}$ whenever $|x_i - x_j| \geq \frac{2L}{\eta_R}\sqrt{\frac{\log(n)}{n}}$ for $L > 4$. The desired inequality (16) is therefore satisfied with probability at least $1 - 4/n$, under the assumptions of the proposition.

• 2^{nd} *Case* $x_j < x_i$. With probability higher than $1 - 4/n$, the following upper bound holds for any $\epsilon \in \{-, +\}$,

$$\begin{aligned}\sum_{k\in C^{\epsilon}_{-(i,j)}(i^{\epsilon})} \epsilon(A_{jk} - A_{ik}) &\leq \sum_{k\in C^{\epsilon}_{-(i,j)}(i^{\epsilon})} \epsilon(F_{jk} - F_{ik}) + \Big| \sum_{k\in C^{\epsilon}_{-(i,j)}(i^{\epsilon})} (E_{ik} - E_{jk}) \Big| \\ &\leq \sum_{k\in C^{\epsilon}_{-(i,j)}(i^{\epsilon})} \epsilon(F_{jk} - F_{ik}) + 4\sqrt{n\log(n)} \ , \end{aligned} \tag{35}$$

where we used Lemma A.3 in the last inequality. For any j satisfying (30), we have $|x_i - x_j| < \rho \leq |x_i - x_{i^{\pm}}|$, so j does not belong to any of the sets $C^{-}_{-(i,j)}(i^{-})$ and $C^{+}_{-(i,j)}(i^{+})$ since they satisfy (32) and (31). In other words, $x_k < x_i, x_j < x'_k$ for all $k \in C^{-}_{-(i,j)}(i^{-})$ and $k' \in C^{+}_{-(i,j)}(i^{+})$, which allows us to use the Robinsonian structure (2) to get

$$\begin{aligned} F_{jk} - F_{ik} > 0 \text{ for all } k \in C^{-}_{-(i,j)}(i^{-}) \ , \\ F_{jk'} - F_{ik'} < 0 \text{ for all } k' \in C^{+}_{-(i,j)}(i^{+}) \ . \end{aligned}$$

(For the case $x_j > x_i$ analyzed here, we actually switch the roles of i, j in (2), since (2) is written for $x_i < x_j$). Therefore, $\sum_{k\in C^{\epsilon}_{-(i,j)}(i^{\epsilon})} \epsilon(F_{jk} - F_{ik}) \leq 0$ for any $\epsilon \in \{-, +\}$, and (35) yields

$$\sum_{k\in C^{\epsilon}_{-(i,j)}(i^{\epsilon})} \epsilon(A_{jk} - A_{ik}) \leq 4\sqrt{n\log(n)} \tag{36}$$

which holds with probability higher than $1 - 4/n$. The right-hand side of (36) is (strictly) smaller than $L\sqrt{n\log(n)}$ for $L > 4$, so, with high probability, (16) is not satisfied and the CE algorithm moves to the next step (the point 4 of the pseudo-code) where it will test (17). If (17) holds, then the algorithm will recover the correct value $h(i, j)$. We can check that (17) holds w.h.p. by following the same analysis as in the 1^{st} case above. (For the current case $x_j < x_i$ however we would switch the indices i and j in the Robinson property (5) to have

$$\begin{aligned} \sum_{k:\ x_k \leq x_j - \rho'} F_{jk} - F_{ik} \geq \eta_R |x_i - x_j| n, \quad \text{if } \ x_j \geq 1/2 - \rho/2 \\ \sum_{k:\ x_k \geq x_i + \rho'} F_{ik} - F_{jk} \geq \eta_R |x_i - x_j| n, \quad \text{if } \ x_i \leq 1/2 + \rho/2 \ . \qquad) \end{aligned}$$

This concludes the analysis of the 2^{nd} case, as well as the proof of Proposition 3.5. □

A.6 Proof of Lemma 3.6 and 3.8

Proof of Lemma 3.6. Throughout the proof, we assume that $x_1 < x_2 < \ldots < x_n$ with no loss of generality, and we assume a direction for h, say $h(1,n) = -1$. Denote by h^* the correct comparison function defined by $h^*(s,t) = 1 - 2\mathbb{1}_{x_s < x_t}$. Let π_h be the permutation induced by h, as defined in (18).

Let j be an index in $[n]$ such that $x_j \geq x_i + 2\nu + \eta_s$. In order to prove that $\pi_h(j) > \pi_j(i)$, it is enough to show that

$$S_h(j) - S_h(i) := \sum_{k=1}^{n} (h(j,k) - h(i,k)) \geq 0 \quad \text{and } h(i,j) = -1 \ . \tag{37}$$

In fact, checking $h(i,j) = -1$ is only useful in the case of a tie $S_h(j) - S_h(i) = 0$.

We introduce a partition of the latent space $[0,1]$, composed of five consecutive intervals:

$$\begin{aligned} &I_1 = [0, x_i - \nu], \quad I_2 = (x_i - \nu, x_i + \nu), \quad I_3 = [x_i + \nu, x_j - \nu], \\ &I_4 = (x_j - \nu, x_j + \nu), \quad I_5 = [x_j + \nu, 1], \end{aligned}$$

assuming that $x_i > \nu$ and $x_j + \nu < 1$ $\big($the other cases $x_i \leq \nu$ or $x_j + \nu \geq 1$ can be analyzed similarly, with a slight adaptation of the current proof$\big)$. This spatial partition is associated with the following partition of indices: $R_s = \{k \in [n] : x_k \in I_s\}$, $s \in [5]$.

For $x_k \in I_1 \cup I_5$, we have $|x_i - x_k| \wedge |x_j - x_k| \geq \nu$. Then, since h has an error less than ν, the values $h(i,k)$ and $h(j,k)$ are equal to the values $h^*(i,k)$ and $h^*(j,k)$ of the correct comparison function h^*. Hence, $h(i,k) = h(j,k) = 1$ for all $k \in R_1$, and $h(i,k) = h(j,k) = -1$ for all $k \in R_5$, thus leading to

$$\sum_{k \in R_1 \cup R_5} (h(j,k) - h(i,k)) = 0 \ .$$

For $x_k \in I_2$, we have $x_j - x_k \geq \nu$, hence $h(j,k)$ matches the correct value $h^*(j,k) = 1$ for all $k \in R_2$. Since h only takes values in $\{-1, 0, 1\}$, we have $|h(i,k)| \leq 1$ for all k, so that

$$\sum_{k \in R_2} (h(j,k) - h(i,k)) \geq 0 \ .$$

By symmetry for $x_k \in I_4$, a similar reasoning yields $\sum_{k \in R_4} (h(j,k) - h(i,k)) \geq 0$.

For $x_k \in I_3$, we have $|x_i - x_k| \wedge |x_j - x_k| \geq \nu$, so $h(i,k)$ and $h(j,k)$ match the true values $h^*(i,k)$ and $h^*(j,k)$, which means that $h(i,k) = -1$ and $h(j,k) = +1$. Hence

$$\sum_{k \in R_3} (h(j,k) - h(i,k)) \geq 2 \, \# R_3 \ ,$$

where $\# R_3$ is the cardinal number of the set R_3.

Gathering the last displays together, we obtain that the desired inequality (37) holds. This concludes the proof of Lemma 3.6. □

Proof of Lemma 3.8. Let h be a comparison function with an error less than ν, and, with no loss of generality, assume that the direction of h is $h(1,n) = -1$, with $x_1 < x_2 < \ldots < x_n$.

Denote by h^* the correct comparison function $h^*(s,t) = 1 - 2\mathbb{1}_{x_s<x_t}$. Given the output π_{ref} of the PE algorithm (described in section 3.3), and indices $i, j \in [n]$ satisfying $2\nu \leq x_j - x_i$, we want to show that $\pi_{ref}(i) < \pi_{ref}(j)$.

• 1st *Case* $g_h(i) \neq g_h(j)$. Following the lines of the proof of Lemma 3.6, we know that $g_h(i) \leq g_h(j)$. Hence, $g_h(i) < g_h(j)$, leading to $\pi_{ref}(i) < \pi_{ref}(j)$.

• 2nd *Case* $g_h(i) = g_h(j)$. We denote by k the number $k := g_h(i) = g_h(j)$. If $C_k = 2$, then $\pi_{ref}(i) = k - 1 < k = \pi_{ref}(j)$. Assume therefore that $C_k \geq 3$ in the sequel.

The PE algorithm calls to the DS algorithm, which will call itself using sets E with smaller and smaller cardinal numbers $\#E$. Hence, the procedure will end in a finite number of calls to the DS algorithm. In each of these calls, the DS algorithm computes $E_s^-, s, E_s^+ = \mathrm{SIT}(h, E)$. We separate the analysis of this output in two cases.

2nd.(a) Let $s \in \{i, j\}$. We have $h(i,j) = h^*(i,j) = -1$ and $h(j,i) = h^*(j,i) = 1$ since the comparison function h has an error less than ν, and i,j satisfy $2\nu \leq x_j - x_i$. Hence, the SIT algorithm orders i, j correctly, which means that $j \in E_s^+$ when $s = i$, and $i \in E_s^-$ when $s = j$. In any case, the output π_{ref} of the PE algorithm satisfies $\pi_{ref}(i) < \pi_{ref}(j)$.

2nd.(b) Otherwise, when $s \notin \{i, j\}$, we have $i, j \in E_s^- \cup E_s^+$. If i, j belong to the same set E_s^- or E_s^+, then $\#E_s^- \geq 2$ or $\#E_s^+ \geq 2$, and the DS algorithm will call itself via $\mathrm{DS}(h, E_s^-)$ or $\mathrm{DS}(h, E_s^+)$. Since the cardinal numbers of E_s^- and E_s^+ decrease at each call, we can consider the call having entries $\#E_s^- \leq 1$ and $\#E_s^+ \leq 1$. Thus, i, j do not belong to the same set E_s^- or E_s^+. Depending on the position x_s, we have the following analysis.
· If $x_s \notin (x_i, x_j)$, for instance $x_s < x_i$, then $x_j - x_s \geq 2\nu$. This yields $h(s,j) = h^*(s,j) = -1$ and so $j \in E_s^+$. Since $\#E_s^+ \leq 1$, we necessarily have $i \in E_s^-$, which allows us to conclude that $\pi_{ref}(i) < \pi_{ref}(j)$.
· If $x_s \in (x_i, x_j)$, then $h(s,i) = h^*(s,i) = 1$ or $h(s,j) = h^*(s,j) = -1$. Indeed, h has an error less than ν, and it follows from $|x_i - x_j| \geq 2\nu$ that $|x_i - x_s| \geq \nu$ or $|x_j - x_s| \geq \nu$. Hence, $i \in E_s^-$ or $j \in E_s^+$. Since $\#E_s^- \vee \#E_s^+ \leq 1$, we get that $i \in E_s^-$ and $j \in E_s^+$, leading once again to $\pi_{ref}(i) < \pi_{ref}(j)$.

The proof of Lemma 3.8 is complete. □

B Sketch of proof of Corollary 2.8

As discussed in section 2.1.1, the bi-Lipschitz condition $f \in \mathcal{BL}[\tilde{\alpha}, \tilde{\beta}]$ implies the hypotheses (2), (4) and (5) for some numerical constant $\eta_R > 0$ only depending on $\tilde{\alpha}$. Besides, the x_i's fulfill the spreading condition (6) for $\eta_s \leq C' \left(\sqrt{\log(n)/n}\right)$ with probability at least $1 - 1/n$ for some numerical constant $C' > 0$. Hence, we can apply Corollary 3.7 to get an estimation of the latent ordering with an error smaller than $\epsilon_n = (20\eta_R^{-1} + C')\sqrt{\log(n)/n}$.

To deduce an error of localization of the x_i's, we can rely on Dvoretzky–Kiefer–Wolfowitz (DKW) inequality. Indeed, the $x_1, \ldots, x_n$ are independent and uniformly distributed on $[0,1]$. We shorten $x_1, \ldots, x_n$ by $\mathbf{x}$, and for any interval I we denote $|I|$ its length, and $N_I(\mathbf{x})$

the number of points x_i which lie in I. We deduce from DKW inequality that, for any $t > 0$,

$$\mathbb{P}\left(\sup_{I\subset[0,1]}\left|\frac{N_I(\mathbf{x})}{n} - |I|\right| > 4t\right) \leq 2e^{-2nt^2} \ ,$$

so that, choosing $t = \sqrt{\log(n)/n}$, we obtain

$$\mathbb{P}\left(\sup_{I\subset I}\left|\frac{N_I(\mathbf{x})}{n} - |I|\right| > 4\sqrt{\frac{\log(n)}{n}}\right) \leq \frac{2}{n^2} \ .$$

We can now easily deduce that $\max_{i,j\in[n]}|\hat{x}_i - x_i| \wedge \max_{i,j\in[n]}|\hat{x}_i - (1-x_i)| \leq C_{\tilde{\alpha}}\sqrt{\frac{\log(n)}{n}}$, for some numerical constant $C_{\tilde{\alpha}}$ depending only on $\tilde{\alpha}$. Corollary 2.8 follows. □.

C Proof of Theorem 4.1 (Lower Bound)

Proof of Theorem 4.1. We establish the lower bound $\eta_R^{-1}\sqrt{\log(n)/n}$ in the particular setting where the observations A_{ij} are independent Bernoulli random variables of parameters $F_{ij} = f_0(x_i, x_j)$, for the specific function

$$f_0(x_i, x_j) = \frac{3}{4} - \frac{\tilde{\eta_R}|x_i - x_j|}{2} \ , \tag{38}$$

with $\eta_R = c\tilde{\eta_R}$ for some numerical constant $c \in (0,1)$, and $\tilde{\eta_R} \in [C_0\sqrt{\log(n)/n}, 1]$ where C_0 is a numerical constant that will be set later.

Our minimax lower bound is based on Fano's method as stated below. For two configuration $\mathbf{x}$ and $\mathbf{x}'$ in $\boldsymbol{X}_n$, we denote the Kullback-Leibler divergence of $\mathbb{P}_{(\mathbf{x},f_0)}$ and $\mathbb{P}_{(\mathbf{x}',f_0)}$ by $KL(\mathbb{P}_{(\mathbf{x},f_0)} \,\|\, \mathbb{P}_{(\mathbf{x}',f_0)})$. Given the pseudo-metric D defined in (19), a radius $\epsilon > 0$ and a subset $\mathcal{S} \subset \boldsymbol{X}_n \cong \boldsymbol{\Pi}_n$, the packing number $\mathcal{M}(\epsilon, \mathcal{S}, D)$ is defined as the largest number of points in $\mathcal{S}$ that are at least ϵ away from each other with respect to D. Below, we state a specific version of Fano's lemma.

Lemma C.1 (from [Yu, 1997])**.** *Consider any subset $\mathcal{S} \subset \boldsymbol{X}_n \cong \boldsymbol{\Pi}_n$. Define the Kullback-Leibler diameter of $\mathcal{S}$ by*

$$d_{KL}(\mathcal{S}) = \sup_{\mathbf{x},\mathbf{x}'\in\mathcal{S}} KL(\mathbb{P}_{(\mathbf{x},f_0)} \,\|\, \mathbb{P}_{(\mathbf{x}',f_0)}) \ .$$

Then, for any estimator $\hat{\pi}$ and for any $\epsilon > 0$, we have

$$\sup_{\mathbf{x}\in\mathcal{S}} \ \mathbb{P}_{(\mathbf{x},f_0)}\left[D(\hat{\pi}, \pi_{\mathbf{x}}) \geq \frac{\epsilon}{2}\right] \geq 1 - \frac{d_{KL}(\mathcal{S}) + \log(2)}{\log \mathcal{M}(\epsilon, \mathcal{S}, D)} \ .$$

In view of the above proposition, we mainly have to choose a suitable subset $\mathcal{S}$, control its Kullback-Leibler diameter, and get a sharp lower bound of its packing number. A difficulty stems from the fact that the loss $D(\hat{\pi}, \pi_{\mathbf{x}})$ is invariant when reversing the ordering of $\pi_{\mathbf{x}}$.

Let $k := (C_1/\tilde{\eta_R})\sqrt{n\log(n)}$, for a small enough numerical constant $C_1 \in (0,1]$ that will be set later. To ensure that $k \leq n/4$, we enforce the condition $\tilde{\eta_R} \geq C_0\sqrt{\log(n)/n}$, with $C_0 := 4C_1$. We introduce $n/4$ vectors $\mathbf{x}^{(s)} \in \boldsymbol{X}_n$, $s = 1, \ldots, n/4$. For each $s \in [n/4]$, let $x_j^{(s)}$ be such that

$$x_j^{(s)} = \frac{j}{n} \ , \forall j \in [n] \setminus \{s, s+k\} \ , \qquad x_s^{(s)} = \frac{s+k}{n}, \quad x_{s+k}^{(s)} = \frac{s}{n} \ .$$

Each vector of positions $\mathbf{x}^{(s)}$ is therefore equal to the vector $(j/n)_{j\in[n]}$ up to an exchange of the two positions s/n and $(s+k)/n$. This collection of $n/4$ vectors is denoted by $\mathcal{S} := \{\mathbf{x}^{(1)}, \ldots, \mathbf{x}^{(n/4)}\}$. Obviously $\mathcal{S} \subset \boldsymbol{X}_n \cong \boldsymbol{\Pi}_n$, and one can readily checks that

$$D(\pi_{\mathbf{x}^{(t)}}, \pi_{\mathbf{x}^{(s)}}) \geq \frac{k}{n} \ , \qquad \forall s, t \in \left[\frac{n}{4}\right], \, s \neq t \ , \tag{39}$$

which in turn ensures that the packing number $\mathcal{M}(\epsilon_n, \mathcal{S}, D)$ of radius $\epsilon_n := k/n$ satisfies $\mathcal{M}(\epsilon_n, \mathcal{S}, D) \geq n/4$.

To upper bound the KL diameter of $\mathcal{S}$, we use the following claim whose proof is postponed to the end of the section.

Claim C.2. *For any $\mathbf{x}, \mathbf{x}' \in \boldsymbol{X}_n$, we have $KL(\mathbb{P}_{(\mathbf{x},f_0)} \,\|\, \mathbb{P}_{(\mathbf{x}',f_0)}) \leq 8\sum_{i,j}(f_0(x_i,x_j) - f_0(x_i',x_j'))^2$.*

Together with the definition (38) of f_0, we get

$$KL(\mathbb{P}_{(\mathbf{x}^{(t)},f_0)} \,\|\, \mathbb{P}_{(\mathbf{x}^{(s)},f_0)}) \leq C_2 n(\tilde{\eta_R}\epsilon_n)^2 \leq C_2(C_1)^2\log(n) \ ,$$

for some numerical constant C_2. Then, choosing the constant C_1 in the definition of k such that $C_1 = (2\sqrt{C_2})^{-1}$ leads to $d_{KL}(\mathcal{S}) \leq \log(n)/4$.

Applying Lemma C.1 to this set $\mathcal{S}$, we arrive at

$$\inf_{\hat{\mathbf{x}}} \sup_{\mathbf{x}\in\mathcal{S}} \quad \mathbb{P}_{(\mathbf{x},f_0)}\left[D(\hat{\pi}, \pi_{\mathbf{x}}) \geq \frac{\epsilon_n}{2}\right] \geq 1 - \frac{\log(n)/4 + \log(2)}{\log(n/4)} \geq \frac{1}{2} \ ,$$

as soon as n is greater than some numerical constant. The lower bound $\epsilon_n/2$ is of the order of $\eta_R^{-1}\sqrt{\log(n)/n}$ since $\tilde{\eta_R}$ is equal to η_R up to a multiplicative numerical constant. Theorem 4.1 is proved. □

Proof of Claim C.2. By definition of the Kullback-Leibler divergence, and $F_{ij} := f_0(x_i, x_j)$ and $F'_{ij} := f_0(x'_i, x'_j)$, we have

$$KL(\mathbb{P}_{(\mathbf{x},f_0)} \,\|\, \mathbb{P}_{(\mathbf{x}',f_0)}) = \sum_{i<j} F_{ij}\log\frac{F_{ij}}{F'_{ij}} + (1-F_{ij})\log\frac{1-F_{ij}}{1-F'_{ij}} \ ,$$

and since $\log(t) \leq t-1$ for all $t > 0$, it follows that

$$KL(\mathbb{P}_{(\mathbf{x},f_0)} \,\|\, \mathbb{P}_{(\mathbf{x}',f_0)}) \leq \sum_{ij} \frac{(F_{ij}-F'_{ij})^2}{F'_{ij}(1-F'_{ij})} \leq 8\sum_{i,j}(F_{ij}-F'_{ij})^2 \ ,$$

where the second inequality follows from the fact that $1/4 \leq F'_{ij} \leq 3/4$. □

References

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

[Barnard et al., 1995] Barnard, S. T., Pothen, A., and Simon, H. (1995). A spectral algorithm for envelope reduction of sparse matrices. Numerical linear algebra with applications, 2(4):317–334.

[Bradley and Terry, 1952] Bradley, R. A. and Terry, M. E. (1952). Rank analysis of incomplete block designs: I. the method of paired comparisons. Biometrika, 39(3/4):324–345.

[Cai and Ma, 2022] Cai, T. T. and Ma, R. (2022). Matrix reordering for noisy disordered matrices: Optimality and computationally efficient algorithms. arXiv preprint arXiv:2201.06438.

[De Castro et al., 2017] De Castro, Y., Lacour, C., and Ngoc, T. M. P. (2017). Adaptive estimation of nonparametric geometric graphs. arXiv preprint arXiv:1708.02107.

[Diaconis and Janson, 2007] Diaconis, P. and Janson, S. (2007). Graph limits and exchangeable random graphs. arXiv preprint arXiv:0712.2749.

[Diaz et al., 2020] Diaz, J., McDiarmid, C., and Mitsche, D. (2020). Learning random points from geometric graphs or orderings. Random Structures & Algorithms, 57(2):339–370.

[Elson et al., 2004] Elson, J., Karp, R. M., Papadimitriou, C. H., and Shenker, S. (2004). Global synchronization in sensornets. In Farach-Colton, M., editor, LATIN 2004: Theoretical Informatics, pages 609–624, Berlin, Heidelberg. Springer Berlin Heidelberg.

[Fogel et al., 2013] Fogel, F., Jenatton, R., Bach, F., and d'Aspremont, A. (2013). Convex relaxations for permutation problems. In Advances in Neural Information Processing Systems, pages 1016–1024.

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

[Garriga et al., 2011] Garriga, G. C., Junttila, E., and Mannila, H. (2011). Banded structure in binary matrices. Knowledge and information systems, 28(1):197–226.

[Gilbert, 1961] Gilbert, E. N. (1961). Random plane networks. Journal of the Society for Industrial and Applied Mathematics, 9(4):533–543.

[Giraud et al., 2021] Giraud, C., Issartel, Y., and Verzelen, N. (2021). Localization in 1d nonparametric latent space models from pairwise affinities. arXiv preprint arXiv:2108.03098.

[Giridhar and Kumar, 2006] Giridhar, A. and Kumar, P. R. (2006). Distributed clock synchronization over wireless networks: Algorithms and analysis. In Proceedings of the 45th IEEE Conference on Decision and Control, pages 4915–4920.

[Hoff et al., 2002] Hoff, P. D., Raftery, A. E., and Handcock, M. S. (2002). Latent space approaches to social network analysis. Journal of the american Statistical association, 97(460):1090–1098.

[Issartel, 2021] Issartel, Y. (2021). On the estimation of network complexity: Dimension of graphons. Journal of Machine Learning Research, 22(191):1–62.

[Janssen and Smith, 2020] Janssen, J. and Smith, A. (2020). Reconstruction of line-embeddings of graphons. arXiv preprint arXiv:2007.06444.

[Lovász, 2012] Lovász, L. (2012). Large networks and graph limits, volume 60. American Mathematical Soc.

[Natik and Smith, 2021] Natik, A. and Smith, A. (2021). Consistency of spectral seriation. arXiv preprint arXiv:2112.04408.

[Penrose, 2003] Penrose, M. (2003). Random geometric graphs, volume 5. Oxford university press.

[Recanati et al., 2017] Recanati, A., Brüls, T., and d'Aspremont, A. (2017). A spectral algorithm for fast de novo layout of uncorrected long nanopore reads. Bioinformatics, 33(20):3188–3194.

[Rocha et al., 2018] Rocha, I., Janssen, J., and Kalyaniwalla, N. (2018). Recovering the structure of random linear graphs. Linear Algebra and its Applications, 557:234–264.

[Von Luxburg, 2007] Von Luxburg, U. (2007). A tutorial on spectral clustering. Statistics and computing, 17(4):395–416.

[Yu, 1997] Yu, B. (1997). Assouad, Fano, and le Cam. In Festschrift for Lucien Le Cam, pages 423–435. Springer.