

## CONSISTENCY OF SPECTRAL SERIATION

AMINE NATIK<sup>1</sup> AND AARON SMITH<sup>2,3,\*</sup>

**Abstract.** Consider a random graph  $G$  of size  $N$  constructed according to a *graphon*  $w : [0, 1]^2 \mapsto [0, 1]$  as follows. First embed  $N$  vertices  $V = \{v_1, v_2, \dots, v_N\}$  into the interval  $[0, 1]$ , then for each  $i < j$  add an edge between  $v_i, v_j$  with probability  $w(v_i, v_j)$ . Given only the adjacency matrix of the graph, we might expect to be able to approximately reconstruct the permutation  $\sigma$  for which  $v_{\sigma(1)} < \dots < v_{\sigma(N)}$  if  $w$  satisfies the following *linear embedding* property introduced in [Janssen *et al. Electron. J. Statist.* **16** (2022) doi:10.1214/21-EJS1940]: for each  $x$ ,  $w(x, y)$  decreases as  $y$  moves away from  $x$ . For a large and non-parametric family of graphons, we show that (i) the popular spectral seriation algorithm [Atkins *et al., SIAM J. Comput.* **28** (1998) 297–310] provides a consistent estimator  $\hat{\sigma}$  of  $\sigma$ , and (ii) a small amount of post-processing results in an estimate  $\tilde{\sigma}$  that converges to  $\sigma$  at a nearly-optimal rate, both as  $N \rightarrow \infty$ .

**Mathematics Subject Classification.** 62M30, 62G20.

Received July 21, 2025. Accepted December 24, 2025.

### 1. INTRODUCTION

In this paper, we study a statistical version of the *seriation problem*, first introduced by Robinson in [1] as a combinatorial problem. In this problem, we get to observe a graph, and our goal is to order the vertices in such a way that more similar items are placed near to each other. The pioneer of seriation is considered to be Flinders Petrie, an Egyptologist, who called it *sequence dating* in [2]. While the seriation originates from archaeology, it has recent applications to ecology, sociology, biology, and other topics. See [3] for a historical overview on the seriation problem.

We relate the seriation problem from [2] to the notation used in this article. We begin with a collection of  $N \in \mathbb{N}$  objects, with the  $i$ 'th object associated with an unobserved point  $v_i \in [0, 1]$ . In [2], these objects corresponds to a set of graves, with  $v_i$  the age of the  $i$ 'th grave. We then get to observe a symmetric graph  $G = (V, E)$  with vertex set  $V = \{v_1, v_2, \dots, v_N\}$ . Informally, we expect that an edge is more likely to appear between vertices  $v_i$  and  $v_j$  that are closer together. Continuing the example of [2], the observed graph is constructed by putting an edge between  $v_i$  and  $v_j$  if we find the same type of artifacts in both the  $i$ 'th and  $j$ 'th graves. This co-occurrence is more likely when the two graves have similar ages. The *goal* in seriation is then to

---

*Keywords and phrases:* Statistical seriation, permutation learning, spectral embedding, graphon.

<sup>1</sup> Département de Mathématiques et Statistiques Université de Montréal Mila - Quebec Artificial Intelligence Institute 6666 St Urbain St, Montreal, Canada.

<sup>2</sup> Department of Mathematics and Statistics University of Ottawa 585 King Edward Ave, Ottawa, Canada.

<sup>3</sup> Tutte Institute for Mathematics and Computing, Ottawa, Canada.

\* Corresponding author: [asmi28@uottawa.ca](mailto:asmi28@uottawa.ca)

find an ordering  $\sigma : V \mapsto \{1, 2, \dots, N\}$  for which

$$v_{\sigma(1)} < v_{\sigma(2)} < \dots < v_{\sigma(N)}.$$

In [2], this means inferring a chronological ordering of the true ages of the set of graves by observing the different types of archaeological finds on each grave.

Many versions of the seriation problem have been studied. Perhaps the simplest occurs when the observed graph satisfies the following property: for all  $v_i < v_j < v_k$ ,

$$(v_i, v_k) \in E \implies (v_i, v_j) \in E \text{ and } (v_j, v_k) \in E. \tag{1.1}$$

Graphs satisfying this property are called *Robinsonian*, named after Robinson who first introduced this property in [1]. We call this the *noiseless* seriation problem, as there will never be any “missing” edges and a perfect ordering exists.

**Remark 1.1** (Reversals and Recovery). Since we can only observe the graph  $G$ , there is no way to distinguish between an ordering  $\sigma$  and its reversal  $\text{rev}(\sigma)$ , defined by

$$\text{rev}(\sigma)(i) = N + 1 - \sigma(i). \tag{1.2}$$

This means that, although we say that we wish to find an ordering  $\hat{\sigma}$  that is close to the ordering  $\sigma$  that satisfies  $v_{\sigma(1)} < v_{\sigma(2)} < \dots < v_{\sigma(N)}$ , the best that we can hope for is an ordering  $\hat{\sigma}$  that is close to *either*  $\sigma$  *or*  $\text{rev}(\sigma)$ . We expect that this non-identifiability is rarely a problem in practical applications: even a small amount of additional information lets you distinguish between the extremely different estimates  $\hat{\sigma}$  and  $\text{rev}(\hat{\sigma})$ .

To avoid repeatedly discussing this small non-identifiability issue, we usually hide it by the following small abuse of notation. For any metric  $d$  that can be applied to orders, we always intend the symmetrized version

$$d(\sigma_1, \sigma_2) = \min(d(\sigma_1, \sigma_2), d(\text{rev}(\sigma_1), \sigma_2)) \tag{1.3}$$

unless we explicitly say otherwise. We only need to pay attention to this issue in step 3 of our post-processing algorithm (Alg. 3), where we consider several different estimated orderings and must “align” them.

This noiseless version of the seriation problem has been well-studied in the computer science literature, and there are several algorithms that are known to return a correct  $\sigma$  very quickly; we believe that [4] is the fastest. Of the provably-correct algorithms, one of the simplest and best-known is the *spectral seriation algorithm* (see Alg. 1 below) introduced in [5]. This algorithm is based on the spectrum of the *Laplacian* of the observed graph.

In realistic examples, the data is noisy and (1.1) will not hold exactly. This motivates the *statistical seriation problem*, in which the edges of the graph  $G$  are seen as random and (1.1) only holds “on average”. We study the following natural generalization introduced in [6]: for all  $v_i < v_j < v_k$ ,

$$\mathbb{P}((v_i, v_k) \in E) \leq \min(\mathbb{P}((v_i, v_j) \in E), \mathbb{P}((v_j, v_k) \in E)). \tag{1.4}$$

Random graphs satisfying this inequality are known as *Robinsonian*. In this situation we typically can’t recover  $\sigma$  exactly, and so the goal becomes *estimating*  $\sigma$ . Algorithm 1, the spectral seriation algorithm, is commonly applied to the statistical seriation problem as well as the noiseless seriation problem.

We give the usual spectral algorithm below, but first set notation. For any integer  $n \in \mathbb{N}$  define  $[n] = \{1, 2, \dots, n\}$ . For any set  $S$  and any function  $f : S \mapsto \mathbb{R}$ , define the associated ordering  $\sigma_f : S \mapsto [|S|]$  by the formula

$$\sigma_f(i) = |\{j \in S : f(j) \leq f(i)\}|, \tag{1.5}$$

breaking ties in  $f$  arbitrarily to make the resulting function bijective. If  $S = [N]$  and  $f$  does not contain any repeated values, it is clear that  $f(\sigma_f^{-1}(1)) \leq f(\sigma_f^{-1}(2)) \leq \dots \leq f(\sigma_f^{-1}(N))$ .

The spectral seriation algorithm is based on the eigenvalues of a related Laplacian matrix, rather than the adjacency matrix itself. We recall that the (unnormalized) Laplacian matrix of a graph with  $N$  vertices and adjacency matrix  $A = (a_{i,j})_{1 \leq i,j \leq N}$  is given by  $L_A = (D - A)$ , where the *degree matrix*  $D = (d_{ij})_{1 \leq i,j \leq N}$  is the diagonal matrix with entries  $d_{ii} = \sum_{j=1}^N a_{ij}$ . The spectral seriation algorithm introduced in [5] is as follows:

---

**Algorithm 1:** The spectral seriation algorithm.

---

**Input :** A graph  $G$ .

- 1 Compute the adjacency matrix  $A$  of the graph  $G$ .
- 2 Compute the Laplacian matrix  $L_A$ .
- 3 Find the eigenvector  $\phi$  of  $L_A$  with second-smallest eigenvalue, called the *Fiedler vector*.
- 4 Compute the permutation  $\sigma_\phi$  according to equation (1.5).

**Output:** The permutation  $\sigma_\phi$ .

---

As shown in [5], under the satisfying property (1.1) this algorithm always returns a correct permutation. The same algorithm also gives a popular estimator for the *statistical* seriation problem. The main contribution of this paper is the study of the performance of this algorithm for the statistical seriation problem under fairly general conditions, which we now introduce.

A *graphon* is a measurable function  $w : [0, 1]^2 \rightarrow [0, 1]$  which is symmetric (*i.e.*,  $w(x, y) = w(y, x)$ ). See [7] for a general introduction to graphons, and [6] for previous work on graphons in the context of seriation. In this paper, we say that a random graph  $G = (V, E)$  with  $N$  vertices is sampled from a graphon  $w$  and write  $G \sim w$  if: (i) the vertices of the graphs  $V = \{v_1, v_2, \dots, v_N\}$  are identified with their embedding  $v_i = \frac{i}{N} \in [0, 1]$  in a latent space and (ii) the edges of  $G$  are sampled independently with probabilities:<sup>1</sup>

$$\mathbb{P}((v_i, v_j) \in E) = w(v_i, v_j) \tag{1.6}$$

for  $i \neq j$ .

Our paper includes two main results on this problem. We now give weak but-easy-to-state versions of our main theorems, applying to the following family of geometric random graphs with smooth and monotone link functions:

**Definition 1.2** (“Nice” Random Graphs). We say that a graphon  $w$  is “nice” if  $w$  can be written as  $w(x, y) = R(|x - y|)$  for all  $x, y \in [0, 1]$ , where  $R : [0, 1] \rightarrow (0, 1)$  is a  $\mathcal{C}^1$  function such that  $\sup_{x \in [0, 1]} R'(x) < 0$ .

To give simple concrete families of examples: (i) the graphons  $w(x, y) = 1 - p|x - y|^q$  belong to this family when  $0 < p \leq 1$  and  $1 \leq q < \infty$ , and (ii) the graphons  $w(x, y) = \exp\left(\frac{-(x-y)^2}{2\sigma^2}\right)$  for any  $\sigma > 0$  (known as the RBF kernel) also belong to this family.

Denote by  $id_N$  the identity permutation on  $[N]$ , and for all  $\gamma > 1$  and  $N \geq 1$  define  $\varepsilon_\gamma(N) := \exp(-\log^\gamma(N))$ . For  $v \in \mathbb{R}^N$ , denote by  $\|v\|_1 = \sum_j |v_j|$  and  $\|v\|_\infty = \max_j |v(j)|$  the usual  $L^1$  and  $L^\infty$  norms. The following is an immediate consequence of Theorems 2.8 and 2.13, as we show in Section 7:

**Theorem 1.3.** *Let  $w$  be a graphon of the form given in Definition 1.2. Consider a family of random graphs  $\{G^{(N)}\}_{N \in \mathbb{N}}$  where each  $G^{(N)}$  has exactly  $N$  vertices, and is sampled from  $w$  in the sense of equation (1.6) (*i.e.*,  $G^{(N)} \sim w$ ), and let  $\hat{\sigma}^{(N)}$  be the output of Algorithm 1 with input  $G^{(N)}$ . For any fixed  $\varepsilon \in (0, 1)$  and  $\eta \in (0, 1/3)$ ,*

---

<sup>1</sup>In the larger graphon literature, it is more common to view  $v_i$  as being i.i.d. samples taken uniformly from  $[0, 1]$ . The analysis in this paper can be carried out in that case with a few small adjustments (the initial spectral argument loses factors of  $\log(n)$  to ensure that the spacing is not too far from uniform; these losses are eliminated in the cleanup algorithm of Section 6, which already uses subsamples of the vertices that are not exactly evenly spaced). We use this embedding to more closely match previous work on special cases of the seriation problem, such as [8–10].

there exists a positive constant  $C = C(w) < \infty$  and an integer  $N_0 = N_0(w, \varepsilon, \eta)$  so that for all  $N \geq N_0$

$$\mathbb{P} \left( \left\| \hat{\sigma}^{(N)} - id_N \right\|_1 \leq CN^{2-\eta} \right) > 1 - \varepsilon. \tag{1.7}$$

Furthermore, denote by  $\tilde{\sigma}^{(N)}$  the “post-processed” estimate obtained from Algorithm 3 with input  $G^{(N)}$  and parameters  $\alpha = 0.05$ ,  $\beta = 0.31$ . Then for any  $\gamma > 1$  and  $\epsilon > 0$ , there exist a positive constant  $\delta = \delta(w, \gamma, \epsilon)$  and an integer  $N_0 = N_0(w, \gamma, \epsilon)$  so that

$$\mathbb{P} \left( \left\| \tilde{\sigma}^{(N)} - id_N \right\|_\infty \leq \epsilon \sqrt{N \log^\gamma(N)} \right) > 1 - \epsilon_\gamma^\delta(N) \tag{1.8}$$

for all  $N \geq N_0$ .

We note that Inequality (1.8) (which follows from Thm. 2.13) is much stronger than (1.7) (which follows from Thm. 2.8) in at least two ways: the norm in which convergence is measured is much stronger, and the rate of convergence is much faster (even after rescaling to account for the different norms). The question of rate-optimality for this problem turns out to be slightly subtle, and we discuss it at greater length in Section 8. However, we can say that it is not possible to do substantially better than Inequality (1.8) for all reasonable graphons - see [11] for a proof that the bound in Inequality (1.8) is optimal (up to a factor of  $\text{polylog}(N)$ ) for a broad class of graphons that are similar to those studied in this paper.

### 1.1. Previous work and main contributions

Informally, the main results of our paper are that (i) the popular spectral embedding algorithm gives reasonably good results for the seriation problem under reasonable assumptions on the nonparametric graphon model, and (ii) a simple post-processing step can be added to obtain very good results. We view this as important reassurance for practitioners: spectral embedding algorithms are very popular, but there was little theoretical support for using these algorithms to fit anything but a few small parametric families of models. Our results provide evidence for the folk belief that spectral algorithms, with appropriate coarse post-processing, are quite insensitive to the details of the generating process  $w$ .

There are many papers in the literature that have either similar-looking results or similar-looking methods, so we give a quick survey and then summarize the most interesting differences.

#### 1.1.1. Previous work: similar algorithms and proofs, different questions

Spectral embedding algorithms are very popular in many contexts, especially in the *spectral clustering* literature (see *e.g.* [12, 13] and variants such as [14–16]). The idea behind all these algorithms is to calculate the first few eigenvectors of a matrix related to your data matrix, use these to embed your data in a low-dimensional space, then use a popular clustering algorithm such as  $k$ -means to cluster points in this representation. As with the seriation problem, this is known to give exactly the desired answer in simple “noiseless” situations.

In “noisy” situations, the most common analyses follow the approach of [17]. In that paper, the authors show that, under appropriate conditions, the eigenvectors of the observed data matrix will converge to the eigenfunctions of a certain limiting operator. When this convergence happens, the associated spectral embedding will also converge to an embedding related to the limiting operator. One can then bound the difference between the clustering associated with the observed data and the limiting operator, obtaining a consistency result.

The proof of our first main bound, Inequality (1.7), follows essentially this approach. We emphasize two main differences:

1. **Definition of “correct” answer:** In both seriation and spectral clustering, we typically observe a graph constructed from unobserved latent points, and we try to analyze our algorithms by analyzing the convergence of this graph to a limiting operator. One striking difference is that it is common to define the goal of a spectral clustering algorithm in terms of that limiting operator (see *e.g.* results 1 and 2 of

[17]), while in spectral seriation we have an independently-defined goal (see Eq. (1)) and use that limiting operator only as a proof technique.

The main consequence of this difference is that, even after proving convergence to some limiting operator, we must check that the seriation obtained from the limiting operator actually agrees with the desired ordering. Checking this turns out to be non-trivial unless the generating graphon  $w$  is very special, and relating the limiting operator to the embedded positions takes up most of Section 3.

2. **Difficulty of problem:** In spectral clustering, the “correct” clustering maps vertices to a finite set  $\{1, 2, \dots, k\}$ , where  $k$  does not grow with the size  $N$  of the observed graph. In seriation, the “correct” ordering is a permutation on  $\{1, 2, \dots, N\}$ . The fact that the seriation problem attempts to estimate a much higher-dimensional object makes the problem substantially messier in a few ways. For example, it is often possible to perfectly recover the correct clustering with a finite sample size even without precise information about the data-generating process (see *e.g.* [18]), but it is usually not possible to perfectly recover the solution to the seriation problem without having very accurate *a priori* knowledge of the graphon  $w$  (see *e.g.* [19]).

We next relate our second main result, Inequality (1.8), to the existing literature.

Although analysis of the spectral embedding gives useful information, there are many possible spectral embeddings and it is difficult to find and analyze the “best” one (see *e.g.* the discussion of various Laplacians in [17]). An alternative is to use a simple spectral embedding to obtain an initial rough solution, then use a simple post-processing algorithm to refine it. This is known to work well for the stochastic block model, the most popular toy model for the spectral clustering problem [20].

We take a similar approach here to prove Inequality (1.8). The main novelty is that we need to develop a post-processing algorithm, as the algorithms used for spectral clustering are not appropriate. Again, the fact that exact recovery is impossible adds some complication to our analysis.

### 1.1.2. Previous work: similar questions

Our work was inspired by [21, 22], which studied random graphs that were very similar, respectively, to the one-parameter graphon  $w(x, y) = p \mathbf{1}_{|x-y| \leq 0.5}$  for some  $p \in [0, 1]$  and the single graphon  $w(x, y) = 1 - |x - y|$ . Like us, [22] follows in the footsteps of [17] and analyzes the spectral seriation algorithm as a perturbation of a limiting algorithm. The main differences between our paper and [21, 22] are:

1. **Inequality (1.7):** We deal with a large nonparametric family of models, while [21, 22] are both restricted to small parametric families. Furthermore, we allow for sequences of graphs with total degree growing more slowly with  $N$  than those in [22], and we allow more noise than [21].
2. **Inequality (1.8):** Using our post-processing algorithm, we obtain sharper results than previous work (even in the special cases that were focused on).
3. **Technical differences:** Both [21, 22] rely on detailed knowledge of the spectrum of a limiting operator, which is only possible because they study rather special families of graph models. Since we deal with the nonparametric case, we do not have access to this information and need substantially different arguments.

The results in this paper are also closely related to those in [19], which share an author. Both papers study a large nonparametric family of graphons, both papers study convergence in the same pseudometric, and both results apply to the same simple “toy” models. The main difference is that the present paper analyzes the very popular and relatively simple spectral seriation algorithm, while [19] develops new and rather complicated algorithms. As a secondary difference, the assumptions in our papers are not comparable. In particular, [19] requires that the *square* of the adjacency matrix satisfy a weaker version of the Robinsonian property, and (as shown in [19]) this does not always hold even for graphons satisfying Assumption 1.2.

Finally, we note that there is a related literature on detecting and recovering embeddings on the *torus*  $[0, 1]$  rather than the *interval*  $[0, 1]$ , such as [8–10]. These papers cover a wider range of algorithms than ours, including some analyses (*e.g.* of row-correlation) that are quite different from ours. However, they have essentially the

same goal. The largest difference is that, like [21, 22], the papers [8–10] are concerned with parametric families of graph with few parameters, and take advantage of this in developing and analyzing their algorithms.

We note that the independent work [23] covers a very similar problem in a nonparametric setting, and was put on arXiv within a few months of this paper. There are many technical differences that make direct comparisons difficult - for example, our graph models are not quite the same, and their main conclusion concerns the relative error of an estimate of the latent position of the vertices rather than the absolute error of an estimate of their ordering. The assumptions also do not translate precisely. However, the results are qualitatively quite similar: they study graphs sampled from a nonparametric collection of bi-Lipschitz model operators, and show that a quantity closely related to the sup-norm error of the estimated ordering converges at roughly the statistical rate of  $\frac{1}{\sqrt{n}}$ .

### 1.1.3. Previous work: summary and further reading

Our work belongs to a very large literature on spectral embedding methods, and a smaller literature on attempting to order or embed vertices on an interval or cycle. The most important differences are:

1. In contrast to most previous work on related questions, we consider a large nonparametric family of graphons. In particular, while our assumptions appear more technical than theirs, this is largely because we are considering a much broader class of models - they are often straightforward to check for the most common explicit parametric families.
2. In contrast to previous work on spectral clustering, we must check that the embeddings associated with our limiting operators correspond to the original latent positions.
3. In contrast to our own previous work [19], we consider both a large nonparametric families of graphons *and* a simple algorithm based on a widely-used approach.

Finally, we mention recent work on several problems that have a similar feel to seriation, but which are not directly comparable to our note. Several papers, such as [24] and [25], have studied what amount to other relaxation of the condition in equation (1.1). These relaxations are also reasonable versions of the “statistical” or “approximate” seriation problem, though the details are quite different.

Since our preprint was written, the very interesting paper [26] carefully compares the best rates of several reasonable algorithms, showing that in their problem (as we suspect in this problem) spectral seriation is provably far from optimal. While the details of the problem in [26] are very different from the problem studied here (they assume that their graphs have a Toeplitz structure, which is a much larger restriction than the “soft” assumptions in this paper; they consider additive continuous noise rather than sampled discrete edges; *etc*), and so it is not straightforward to directly translate results, we expect qualitatively-similar results to hold for seriation as well. There have also been many articles on the closely-related problem of estimating a permutation from pairwise comparisons (see *e.g.* [27]) or other estimates related to the entries of a random matrix whose expected value is a Robinsonian matrix. See the longer paper [19] for a longer section relating our statistical seriation question to this work.

## 1.2. Paper Guide

In Section 2, we introduce the main pieces of notation used throughout the rest of the paper and formally state Theorems 2.8 and 2.13, stronger versions of the first and second parts of Theorem 1.3.

In Section 3, we check that a conjectured limit of the graph Laplacian has certain good properties - primarily that an associated eigenfunction is monotone. In Section 4, we check that our graph Laplacian really does converge to this conjectured limit in a sufficiently strong sense. In Section 5, we put together the bounds in Sections 3 and 4 to prove Theorem 2.8.

In Section 6, we introduce the new post-processing algorithm and prove Theorem 2.13.

In Section 7, we check that the assumptions of Theorems 2.8 and 2.13 are weaker than the assumptions of Theorem 1.3. Finally, we have a short concluding discussion in Section 8.

## 2. NOTATION AND MAIN RESULTS

We use standard big-O notation when measuring the running times of algorithms. In particular, for two functions  $f, g : \mathbb{N} \mapsto [0, \infty)$ , we say  $f = O(g)$  if there exists a constant  $C$  such that  $f(x) \leq Cg(x)$  for all  $x$ . Similarly, we say  $f = \Theta(g)$  if both  $f = O(g)$  and  $g = O(f)$ .

### 2.1. Notation for random graphs

Let  $w : [0, 1]^2 \rightarrow [0, 1]$  be a graphon. For each integer  $N$ , consider the partition  $(\mathcal{I}_i^N)_{i=1, \dots, N}$  of the interval  $[0, 1]$  defined by  $\mathcal{I}_1^N = [0, 1/N]$  and  $\mathcal{I}_i^N = ((i-1)/N, i/N]$  for  $i = 2, \dots, N$ . For a given graphon  $w$  we define the *model matrix*  $P^{(N)} \in [0, 1]^{N \times N}$  by  $P_{ij}^{(N)} := w(i/N, j/N)$  for all  $i, j \in \{1, 2, \dots, N\}$ .

For each  $N \in \mathbb{N}$ , we consider the sequence of random graphs  $G^{(N)} \sim \rho_N w$ , where  $\rho_N \in (0, 1]$  is a sequence regulating the sparsity of the graph. We denote by  $\widehat{P}^{(N)}$  the adjacency matrix of the graph  $G^{(N)}$ . We recall that, the graph  $G^{(N)}$  has  $N$  vertices and the edges are constructed as follows: for each  $1 \leq i < j \leq N$ , we add an undirected edge between the  $i$ 'th and  $j$ 'th vertices with probability  $\rho_N P_{ij}^{(N)}$ ; these edges are added independently. In other words,

$$\widehat{P}_{ij}^{(N)} \stackrel{\text{ind}}{\sim} \text{Bernoulli} \left( \rho_N w \left( \frac{i}{N}, \frac{j}{N} \right) \right). \quad (2.1)$$

We view the graphon  $\rho_N w$  as the limit of the sampling matrices  $\widehat{P}^{(N)}$  under suitable rescaling (see [28] for an explanation of how this notion of ‘‘graph limit’’ can be made precise). This motivates the graphon analogue to property (1.4): for  $x, y, z \in [0, 1]$ ,

$$\{x < y < z\} \implies \{w(x, z) \leq \min(w(x, y), w(y, z))\}. \quad (2.2)$$

In analogy to (1.4), we call graphons with this property *Robinsonian* (these are also what [6] calls *diagonally increasing*).

We define the sampled graphon  $w_N$  by

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

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

Similarly, the random graphon  $\widehat{w}_N$  is defined by:

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

### 2.2. Review of relevant facts from spectral theory

We next need some basic notation related to operators. We denote by  $L^2([0, 1])$  the usual Hilbert space on (equivalence classes of) functions  $f : [0, 1] \mapsto \mathbb{R}$ , with inner product given by  $\langle f, g \rangle := \int_0^1 f(x)g(x)dx$  and norm denoted  $\|f\|$ . For any function  $f$  or operator  $T$ , denote by  $rg(f)$  (respectively  $rg(T)$ ) the range of the function (respectively operator). Let  $e \in L^2([0, 1])$  denote the constant function,  $e(x) := 1$  for all  $x \in [0, 1]$ . We use blackboard bold symbols, such as  $\mathbb{T}$ , to denote linear operators acting on  $L^2([0, 1])$ . The induced (operator) norm is defined as

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

For a self-adjoint bounded linear operator  $\mathbb{T}$  we denote by  $\sigma(\mathbb{T})$  the spectrum of  $\mathbb{T}$ , which consists of the set of all  $\lambda \in \mathbb{R}$  such that  $\mathbb{T} - \lambda\mathbb{I}$  is not bijective. For any operator  $\mathbb{T}$ , we denote by  $\ker(\mathbb{T})$  the kernel of  $\mathbb{T}$ . We recall some basic facts about the spectrum (see *e.g.* [29]). The spectrum is always non-empty, closed and bounded. We define the *discrete spectrum*  $\sigma_d(\mathbb{T})$  to be the part of  $\sigma(\mathbb{T})$  which consists of all isolated points with finite algebraic multiplicity<sup>2</sup>. If  $\lambda \in \sigma_d(\mathbb{T})$ , then  $\lambda$  is an eigenvalue of  $\mathbb{T}$  - that is there is some non-zero  $f \in L^2([0, 1])$  such that  $\mathbb{T}f = \lambda f$ . We also define the *essential spectrum*  $\sigma_{\text{ess}}(\mathbb{T}) = \sigma(\mathbb{T}) \setminus \sigma_d(\mathbb{T})$ . According to [30] the essential spectrum  $\sigma_{\text{ess}}(\mathbb{T})$  is the set of all numbers  $\lambda$  in the spectrum  $\sigma(\mathbb{T})$  of  $\mathbb{T}$  such that at least one of the following conditions holds:

- (i) the range  $rg(\mathbb{T} - \lambda\mathbb{I})$  of  $\mathbb{T} - \lambda\mathbb{I}$  is not closed;
- (ii)  $\lambda$  is a limit point of  $\sigma(\mathbb{T})$ ;
- (iii)  $\cup_{r=1}^{\infty} \ker((\mathbb{T} - \lambda\mathbb{I})^r)$  is infinite dimensional.

The essential spectrum is always closed, and the discrete spectrum can only have accumulation points on the boundary to the essential spectrum. For more results on the spectrum of linear operator we suggest [29]. Denote by  $\lambda_{\max}(\mathbb{T})$  the spectral radius of  $\mathbb{T}$ ,

$$\lambda_{\max}(\mathbb{T}) = \sup \{|\lambda| : \lambda \in \sigma(\mathbb{T})\}.$$

Note that  $\lambda_{\max}(\mathbb{T}) \leq \|\mathbb{T}\|$ .

### 2.3. Spectral theory of the graphon-laplacian

We define the *graphon-Laplacian operator*  $\mathbb{L}$  for a given graphon  $w$  as

$$\begin{aligned} (\mathbb{L}f)(x) &:= \int_0^1 w(x, y)(f(x) - f(y))dy \\ &= f(x)d(x) - \int_0^1 w(x, y)f(y)dy, \end{aligned} \tag{2.3}$$

where

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

is the degree of  $x$ . The operator  $\mathbb{L}$  for the graphon  $w$  is the analogous of the Laplacian matrix of a graph. Similarly, let  $\mathbb{L}_N$  and  $\widehat{\mathbb{L}}_N$  denote the Laplacian operators associated to graphons  $w_N$  and  $\widehat{w}_N$  respectively, as defined in equation (2.3). Finally, let  $L^{(N)}, \widehat{L}^{(N)}$  denote the Laplacian operators on  $\mathbb{R}^N$  associated to the matrices  $P^{(N)}$  and  $\widehat{P}^{(N)}$  respectively.

We make some remarks on the spectrum of  $\mathbb{L}$ . Note that the operator  $(\mathbb{W}f)(x) := \int_0^1 w(x, y)f(y)dy$  is a Hilbert-Schmidt operator (see *e.g.* [29]), thus  $\mathbb{W}$  is compact. Nevertheless,  $\mathbb{L}$  is generally not compact since it is the sum of the compact operator  $\mathbb{W}$  and the multiplication operator  $f(x) \mapsto d(x)f(x)$ . It is well known that adding a compact operator does not change the essential spectrum (see *e.g.*, [31]), so the essential spectrum of  $\mathbb{L}$  is the same as the essential spectrum of the multiplication operator  $f(x) \mapsto d(x)f(x)$ . The essential spectrum of this operator is known to be the range of the degree function  $d$ , *i.e.*,  $\sigma_{\text{ess}}(\mathbb{L}) = rg(d)$  (see, *e.g.*, [17]). Note that the linear operator  $\mathbb{L}$  is self-adjoint, bounded and positive.

---

<sup>2</sup>Recall that the algebraic multiplicity of  $\lambda$  is the dimension of  $\cup_k \ker((\mathbb{T} - \lambda\mathbb{I})^k)$ .

It is easy to see that  $\mathbb{L}e = 0$ , which means that  $\lambda_1 = 0$  is an eigenvalue of  $\mathbb{L}$ . In fact it is the smallest eigenvalue, as:

$$\langle \mathbb{L}f, f \rangle = \frac{1}{2} \int_0^1 \int_0^1 w(x, y)(f(x) - f(y))^2 dx dy \geq 0. \quad (2.4)$$

If  $\mathbb{L}f = 0$  and  $w(x, y)$  is strictly positive for almost all  $x, y$ , then Inequality (2.4) implies that  $f \in \text{Vect}(e)$ <sup>3</sup>. Thus,  $\lambda_1 = 0$  is a simple eigenvalue. Denote  $\lambda_1 = 0 < \lambda_2 < \lambda_3 < \dots$  the eigenvalues of  $\mathbb{L}$  in increasing order. We call the *Fiedler value* the smallest non-zero eigenvalue (*i.e.*,  $\lambda_2$ ) and any corresponding eigenfunction  $\phi$  a *Fiedler function*.

## 2.4. Statement of main results

We replace the simple but very strong assumptions of Theorem 1.3 by a series of weaker but more complicated assumptions, then state our main results. Although our main results are more general than Theorem 1.3, we don't aim for maximum generality. Instead, we aim for conditions related to familiar quantities where possible, and give remarks on how these assumptions relate to our toy examples and proofs. Although our assumptions are all stated without reference to the eigenvalues and eigenfunctions of our operators, several of the more technical remarks refer to the ‘‘Fiedler’’ eigenfunction. This eigenfunction is closely analogous to the usual ‘‘Fiedler’’ eigenvector of a matrix already defined, and is discussed more carefully in Section 3.

**Assumption 2.1** (Lipschitz graphon). There exists a constant  $K > 0$  such that

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

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

**Assumption 2.2** (Differentiable graphon). The partial derivative  $\partial w(x, \cdot)/\partial x$  both exists and is non-zero almost everywhere.

**Remark 2.3** (Fixing Assumptions 2.1 and 2.2 for Flat Graphons). We note that Assumptions 2.1 and 2.2 are violated for simple graphons such as  $w(x, y) = 0.5\mathbf{1}_{|x-y| \leq 0.5}$ . This is somewhat unsatisfying, as this simple graphon (and other similar ones) were studied in [8, 9, 22].

Fortunately, these problems are not too difficult to fix, and our arguments were written with this in mind. The proof of Theorem 2.8 only uses Assumption 2.2 (i) inside of the proofs of the intermediate results Theorems 3.1 and 3.2, and (ii) at one step in the proof of Lemma 4.1. Similarly, Assumption 2.1 is only used (i) inside of the proofs of the same intermediate results Theorems 3.1 and 3.2, and (ii) at one step in the proof of Lemma 4.1 and Theorem 4.2.

Thus, both the conclusion and the proof of Theorem 2.8 remain correct if we (i) assume the conclusions of Theorems 3.1 and 3.2 directly and (ii) calculate the three integrals where Assumptions 2.1 and 2.2 are used in the proofs of Lemma 4.1 and Theorem 4.2. These turn out to be straightforward calculations for this toy model.

We can give a slightly more generic version of this discussion. Theorem 2.8 remains true as stated if we substitute the following pair of assumptions for Assumption 2.2:

1. There exists a Fiedler function  $\phi \in \mathcal{C}^1([0, 1])$  and  $L > 0$  such that  $\inf_{x \in [0, 1]} \phi'(x) = L$  (this corresponds to Thm. 3.1).
2. The eigenspace corresponding to eigenvalue  $\lambda_2$  is one-dimensional (this corresponds to Thm. 3.2).

**Assumption 2.4.** The derivative  $d'(x)$  exists for all  $x \in [0, 1]$  and the set  $N := \{x \in [0, 1] : d'(x) = 0\}$  is discrete.

<sup>3</sup>We note here that any Robinsonian graphon  $w$  that satisfies Assumption 2.2 will satisfy  $w(x, y) > 0$  for almost all  $x, y$ . For that reason, the calculations in this paragraph will hold wherever Assumption 2.2 holds.

**Remark 2.5** (Non-Triviality of Asm. 2.4). Note that Assumption 2.4 does not imply that  $d$  is monotone or unimodal -  $d(x)$  is allowed to oscillate. In particular, Assumption 2.4 doesn't imply that you can recover a good ordering of the vertices by just "sorting by degree".

**Remark 2.6** (Where is Assumption 2.4 Used?). Unlike the other assumptions in this section, we don't recall having seen assumptions similar to Assumption 2.4 in other analyses of spectral algorithms. It also seems rather strong, since it requires  $d'(x) \neq 0$  almost everywhere. In fact, this assumption is much stronger than what we really require, and our proofs apply as-written to many examples for which  $d^{-1}(x)$  is of large measure. Assumption 2.4 is included large as a simple-to-check sufficient condition.

In our proofs, Assumption 2.4 is only used in a few steps of the proofs of Theorems 3.1 and 3.2. These theorems state that the Fiedler eigenfunction is monotone and the corresponding eigenspace is one-dimensional. Thus, this assumption can be dropped if there is an alternative way to check these conditions (*e.g.* the eigenfunction and eigenspace can be computed explicitly for the toy example, as in [22]; alternatively, perturbation bounds can be used to compare the eigenfunctions to those of an operator for which they can be computed explicitly).

As a consequence, we note that our proof applies as-written to many examples for which  $d$  is *not* nonzero almost-everywhere.

**Assumption 2.7.** Let

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

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

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

Note that  $\inf(\sigma_{\text{ess}}(\mathbb{L})) = \inf_{x \in [0, 1]} d(x)$ . Applying this with the mini-max principle for self-adjoint operators (see [32]), Assumption 2.7 ensures that the Fiedler value satisfies  $\lambda_2 < \inf_{x \in [0, 1]} d(x)$ .

Although these assumptions may seem complicated, especially Assumption 2.7, all "nice" graphons in the sense of Definition 1.2 satisfy these assumptions. For the proof of this result, see Section 7. Furthermore, note that the most complicated assumption, Assumption 2.7, can be vastly simplified for graphons of the form  $w(x, y) = R(|x - y|)$ : it holds under the weaker (and easier to state) assumption that  $R$  is continuous and satisfies  $R(1) < R(0)$  (see [33]). In practice, we suspect that a practical approach to Assumption 2.7 would be to check that it held for a sampled observed graph.

Our main result is the following consistency result for spectral seriation algorithm:

**Theorem 2.8.** Fix  $\tau \in [0, 1)$ ,  $\eta < \frac{1-\tau}{3}$  and  $\varepsilon \in (0, 1)$ . Let  $\rho_N = N^{-\tau}$ . Fix a Robinsonian graphon  $w$  satisfying Assumptions 2.1 to 2.7. Consider a family of random graphs  $\{G^{(N)}\}_{N \in \mathbb{N}}$  where  $G^{(N)} \sim \rho_N w$  according to equation (2.1). Let  $\hat{\sigma}^{(N)}$  be the output of Algorithm 1 with input  $G^{(N)}$ . Then there exists a positive constants  $C = C(w)$  and an integer  $N_0 = N_0(w, \tau, \eta, \varepsilon)$  so that

$$\mathbb{P} \left( \left\| \hat{\sigma}^{(N)} - id_N \right\|_1 \leq CN^{2-\eta} \right) \geq 1 - \varepsilon \tag{2.5}$$

for all  $N \geq N_0$ .

Moreover, for any  $\gamma > 1$  there exists a positive constants  $C' = C'(w)$  and  $N'_0 = N'_0(w, \tau, \eta, \gamma)$  so that

$$\mathbb{P} \left( \left\| \hat{\sigma}^{(N)} - id_N \right\|_1 \leq C'N^{2-\eta} \right) \geq 1 - \varepsilon_\gamma(N) \tag{2.6}$$

for all  $N \geq N'_0$ , where we recall  $\varepsilon_\gamma(N) = \exp(-\log^\gamma(N))$ .

**Remark 2.9** (Implied Constants). The constant  $C(w)$  in Theorem 2.8 is exactly the prefactor  $\left(1 + \frac{C^2}{L^2}\right)$  appearing in Inequality (5.1). Both constants  $C, L$  in this expression depend on the spectral data of an operator associated with  $w$ :  $L$  is the smallest slope of the so-called Fiedler eigenfunction (see Thm. 3.1 for the definition), while the constant  $C$  is the prefactor appearing in equation (4.14) and depends on the differences between the first few eigenvalues of the same limiting operator. Informally,  $C(w)$  is small if the limiting operator (i) has a top eigenfunction that is not close to flat, and (ii) top eigenvalues that are well-separated.

The constants  $N_0, N'_0$  are defined in equation (4.12). Ignoring polylog factors, the maximum row-degree  $\rho_N N$  need to be large compared to (i) the Lipschitz constant  $K$  associated with the graphon, (ii) the inverse of the maximum expected degree, and (iii)  $\log(\epsilon^{-1})$ .

Note that Inequality (2.6) looks very similar to Inequality (2.5) if we choose  $\epsilon = \epsilon_\gamma(N)$ . Although the two are very similar, the result in (2.6) can not be obtained as a consequence of (2.5). In fact, the values of  $N$  for which the first inequality is satisfied depend on  $N_0$  which itself depend on  $\epsilon$ .

With appropriate post-processing and some additional assumptions, we can achieve tighter estimates. For fixed  $0 < \alpha < 0.5$ , define the functions  $\Psi_R, \Psi_L$  by

$$\Psi_R(x) = \Psi_R(x; \alpha) := \int_{1-\alpha}^1 w(x, y) dy \quad \text{and} \quad \Psi_L(x) = \Psi_L(x; \alpha) := \int_0^\alpha w(x, y) dy. \quad (2.7)$$

These functions count the expected number of neighbours in the right- and left-most  $\alpha$  percent of vertices, though we will suppress the dependence on  $\alpha$  in our notation when writing our main algorithms. Assume:

**Assumption 2.10.** There exists some constants  $0 < \alpha < \beta < 1/2$  such that:

1. **Mean distance inequality:**

$$\inf_{x \in [1-\alpha, 1]} \Psi_R(x; \alpha) > \Psi_R(1 - \beta; \alpha), \quad \inf_{x \in [0, \alpha]} \Psi_L(x; \alpha) > \Psi_L(\beta; \alpha), \quad (2.8)$$

and

2. **Distinguishability inequality:** There exists some  $d_1 > 0$  such that

$$\left\{ \min(|x - z|, |y - z|) \geq \frac{\beta - \alpha}{2} \right\} \Rightarrow \{|w(y, z) - w(x, z)| \geq d_1 |x - y|\} \quad (2.9)$$

for  $x, y, z \in [0, 1]$ .

**Remark 2.11** (Relationship to Assumptions in Other Papers). The assumption that  $w$  is bi-Lipschitz condition is the central assumption in [11], and seems to be very convenient in the analysis of seriation problems.

Combining Assumption 2.10 with previous assumptions is quite close to assuming a bi-Lipschitz condition. Assumption 2.1 requires  $w$  to be Lipschitz. The conclusion of condition (2.9) is quite close to requiring  $w$  to be bi-Lipschitz, except that the inequality on the right-hand side of condition (2.9) only needs to hold for  $x, y, z$  that are sufficiently distant.

**Remark 2.12** (Non-Examples for Assumption 2.10, Part 1). There are many graphons that don't satisfy Part 1 of Assumption 2.10. One very simple family of examples is  $w_C(x, y) = \mathbf{1}_{x, y > C}$  for  $C > 0.5$ . This suggests a much broader construction of “bad” examples: take your favourite Robinson graphon  $w$ , then decrease  $w(x, y)$  for all  $x, y$  small.

This assumption holds also for all graphons that are “nice” in the sense of Definition 1.2. For the proof of this result, see Section 7. When a graphon satisfies Inequality (2.8) for *some* pair  $(\alpha, \beta)$  and also satisfies our other assumptions, it is possible to quickly find the pair  $(\alpha, \beta)$ ; see Remark 6.8.

The following theorem gives a much faster convergence rate under Assumption 2.10:

**Theorem 2.13.** *Let  $\rho_N = 1$ , fix a Robinsonian graphon  $w$  satisfying Assumptions 2.1 to 2.10. Consider a family of random graphs  $\{G^{(N)}\}_{N \in \mathbb{N}}$  where  $G^{(N)} \sim w$  according to equation (2.1). Let  $\hat{\sigma}^{(N)}$  be the output of Algorithm 3 with input  $G^{(N)}$  and parameters  $\alpha, \beta$  as in assumption 2.10. Then for any  $\gamma > 1$  and  $\epsilon > 0$  there exists a positive constant  $\delta = \delta(w, \gamma, \epsilon)$  and an integer  $N_0 = N_0(w, \gamma, \epsilon)$  so that*

$$\mathbb{P} \left( \left\| \hat{\sigma}^{(N)} - id_N \right\|_{\infty} \leq \epsilon \sqrt{N \log^{\gamma}(N)} \right) \geq 1 - \epsilon_{\gamma}^{\delta}(N)$$

for all  $N \geq N_0$ , where we recall  $\epsilon_{\gamma}(N) = \exp(-\log^{\gamma}(N))$ .

**Remark 2.14** (Replacing  $\Psi_R, \Psi_L$  With Other Test Functions ). A critical step of Algorithm 3 is constructing good estimators of the functions  $\Psi_R, \Psi_L$ . Although we give specific functions  $\Psi_R, \Psi_L$  in equation (2.7) and specific estimators in Algorithm 3, our proofs only rely on three properties of this pair of functions: they (i) satisfy Assumption 2.10, (ii) have derivatives bounded away from 0 on  $[0, 1 - \beta]$  and  $[\beta, 1]$  respectively, and (iii) can be estimated with error roughly  $O\left(\sqrt{\frac{\log(N)}{N}}\right)$  from a graph of size  $N$ . We state our results in terms of the specific functions  $\Psi_R, \Psi_L$  because it is easy to (a) compute them and (b) verify our assumptions for the most popular graphon families. However, different comparison functions may be more efficient for other families of graphons, such as the “bad” examples in Remark 2.12, in much the same way that different normalized Laplacians are used in spectral clustering.

## 2.5. Qualitative descriptions of why assumptions are required

In remarks in Section 2.4, we discussed examples for which our main assumptions failed to hold. Here, we briefly explain why these assumptions are necessary - either necessary for the conclusions to hold, or merely necessary for our proofs.

1. **Assumption 2.1:** This assumption is useful for making our proofs “tidier,” and can be substantially weakened at the cost of slightly messier arguments and theorem statements.

Throughout the proofs, we attempt to estimate functions with continuous domains (such as the average degree function  $d(x) = \int w(x, y)dy$ ) from discrete observations. In general, there is no real hope of estimating functions on continuous domains unless the functions vary smoothly (or have some other very special function, such as monotonicity). Assumption 2.1 is a simple-to-state assumption that guarantees these derived quantities all vary smoothly, and thus can be estimated accurately from a finite amount of data. There are many possible substitutes for Assumption 2.1 - most obviously, we expect that being only Lipschitz in “most” places is good enough (see Rem. 2.3).

2. **Assumption 2.2:** This assumption can be slightly weakened for some examples, but our conclusions are in fact false if this assumption is badly violated.

This assumption is used to ensure that the Fiedler eigenfunction is strictly monotone and that the corresponding eigenspace is one-dimensional. It is also used to ensure identifiability of the correct ordering. We can check that both strict monotonicity and identifiability can easily fail by considering the simple Robinson graphon  $w(x, y) = \mathbf{1}_{|x-y| < 0.9}$ : the Fiedler eigenfunction is flat in the middle of its domain, and of course  $w(x, \cdot) \equiv 1$  for  $x \in [0.1, 0.9]$ , so the true ordering is not identifiable from a sampled graph.

3. **Assumption 2.4:** This technical assumption is critical to the proof of one of our lemmas. We believe that it can be substantially weakened.

This assumption is used primarily to check that the Fiedler eigenfunction is  $\mathcal{C}^1$  outside of a discrete set, which is necessary to check that it is strictly monotone. While it is not difficult to construct a graphon whose Fiedler eigenfunction is *not*  $\mathcal{C}^1$  in this way, we are not aware of any graphons that are actually used as models for which this property fails. As a consequence, we expect that this technical condition could be weakened.

4. **Assumption 2.7:** This technical assumption is critical to the proof of one of our lemmas. We believe that it can be slightly weakened.

This assumption is primarily used to ensure that the Fiedler eigenvalue is isolated, which is required in order to use standard perturbation results. While it does seem necessary to have an isolated Fiedler eigenvalue, in principle this isolation could be checked more directly.

5. **Assumption 2.10:** These assumptions could be weakened, but they would require corresponding changes in Algorithm 3.

Informally, the main idea in Algorithm 3 is to estimate certain test functions  $\Psi$  with the property that, for all  $x, y$  in a particular interval  $I$ ,  $\Psi(x) > \Psi(y)$  if and only if  $x > y$ . For the specific intervals  $I_L, I_R$  and test functions  $\Psi_R, \Psi_L$  used in Algorithm 3, Part 1 of Assumption 2.10 ensures that the test functions have our desired properties on the desired intervals and Part 2 ensures that there is enough signal in a random sample to accurately estimate the test functions.

These assumptions are close to necessary for the specific test functions and intervals we use. If Part 1 failed, an incorrect comparison could be made and this could produce an arbitrarily-bad estimated ordering. If Part 2 failed, the noise in the estimate  $\hat{\Phi}_R(x)$  would be large compared to the signal  $|\Phi_R(x) - \Phi_R(y)|$ , and this would again lead to an incorrect comparison. Using different test functions would lead to different conditions.

### 3. PROPERTIES OF THE FIEDLER VALUE AND THE FIEDLER FUNCTION

In this section we establish basic facts about the spectrum of the graphon-Laplacian operator  $\mathbb{L}$ . Throughout this section, we fix a graphon  $w$  satisfying Assumptions 2.1 to 2.7, then define  $\mathbb{L}$  as in equation (2.3). We also denote by  $\lambda_2$  its Fiedler eigenvalue. For  $x \in \mathbb{R}$ , define  $D_x = d^{-1}(\{x\})$ .

The only results we need from this section are the following two theorems, whose proofs will take the rest of the section. The remainder of this section can be safely skipped on a first read of the paper.

**Theorem 3.1.** *There exists a Fiedler function  $\phi \in \mathcal{C}^1([0, 1])$  and  $L > 0$  such that  $\inf_{x \in [0, 1]} \phi'(x) = L$ .*

**Theorem 3.2.** *The eigenspace corresponding to eigenvalue  $\lambda_2$  is one-dimensional.*

The following lemma ensures that most of the eigenfunctions of  $\mathbb{L}$  are  $\mathcal{C}^1$ :

**Lemma 3.3.** *Let  $\lambda$  be any eigenvalue of  $\mathbb{L}$  and  $f$  any eigenfunction associated to the eigenvalue  $\lambda$ . Then the set  $D_\lambda$  is discrete and  $f$  is  $\mathcal{C}^1$  in  $D_\lambda^c$ .*

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

*Proof.* First we show that under Assumption 2.4 the set  $D_\lambda$  is discrete. Let  $N = \{x \in [0, 1] : d'(x) = 0\}$  as in Assumption 2.4. Since  $N$  is discrete, it is enough to show that  $D' = D_\lambda \setminus N$  is discrete. Let  $x \in D'$ , so that  $d(x) = \lambda$  and  $d'(x) \neq 0$ . Moreover,  $d(x+h) - \lambda = d(x+h) - d(x) = d'(x)h + o(h)$ , hence  $d(y) \neq \lambda$  for all  $y \neq x$  in some neighborhood  $U_x$  of  $x$ . Consequently every point  $x \in D'$  has a neighborhood  $U_x$  such that  $D' \cap U_x = \{x\}$  as required.

Now take  $x \in D_\lambda^c$  and  $\Delta > 0$  such that  $x + \delta \notin D_\lambda$  for all  $\delta \in [0, \Delta]$ . Recalling the definition of the operator  $\mathbb{L}$  from equation (2.3),

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

Plugging  $x$  and  $x + \delta$  in the previous equation and taking the difference yields, for  $\delta \in [0, \Delta]$  sufficiently small,

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

For the integral convergence, we apply the dominated convergence theorem. This is justified by the fact that, by Assumption 2.2, the partial derivative exists almost everywhere, and by Assumption 2.1, we have that  $(w(x + \delta, y) - w(x, y))/\delta$  is uniformly bounded by  $K < \infty$ . Therefore  $f$  is differentiable in  $D_\lambda^c$  and for all  $x \in D_\lambda^c$

$$f'(x) = \frac{1}{d(x) - \lambda} \left( \int_0^1 (f(y) - f(x)) \frac{\partial w(x, y)}{\partial x} \right) dy. \quad (3.1)$$

By a similar argument we can show that  $f'(x + \delta) \rightarrow f'(x)$  as  $\delta \rightarrow 0$ , consequently  $f$  is  $\mathcal{C}^1$  on  $D_\lambda^c$ .  $\square$

To prove Theorem 3.1, we introduce the linear operator

$$(\mathbb{M}f)(x) = d(x)f(x) - \int_0^1 \int_0^1 \left( f(z) \frac{\partial w(x, y)}{\partial x} \mathbf{1}_{\{x \leq z \leq y\}} \right) dz dy. \quad (3.2)$$

Note that similarly to  $\mathbb{L}$ , the operator  $\mathbb{M}$  is a self-adjoint, positive and bounded operator. Moreover, it is the sum of a multiplication operator and a compact operator, which means that  $\sigma_{\text{ess}}(\mathbb{M}) = rg(d)$ . We will show that its spectral data is closely related to that of  $\mathbb{L}$ :

**Lemma 3.4.** *The eigenvalues of  $\mathbb{M}$  are  $\lambda_2 < \lambda_3 < \dots$ , that is the spectrum of  $\mathbb{M}$  satisfies  $\sigma(\mathbb{M}) = \sigma(\mathbb{L}) \setminus \{0\}$ . Moreover, for all  $\lambda \neq 0$ ,  $(\lambda, f)$  is an eigenvalue-eigenfunction pair of  $\mathbb{L}$  if and only if there exists a function  $g$  so that  $f'(x) = g(x)$  for all  $x \in D_\lambda^c$  and  $(\lambda, g)$  is an eigenvalue-eigenfunction pair of  $\mathbb{M}$ <sup>4</sup>.*

*Proof.* Let  $(\lambda, f)$  be an eigenvalue-eigenfunction pair of  $\mathbb{L}$  for which  $\lambda \neq 0$ . By Lemma 3.3,  $f$  is differentiable at all  $x \in D_\lambda^c$ , and by Assumption 2.2, the partial derivatives of  $w$  exist almost everywhere. This justifies the calculation

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

Conversely, let  $(\lambda, g)$  be an eigenvalue-eigenfunction of  $\mathbb{M}$ . Then for all  $x \notin D_\lambda$ , straightforward algebra gives

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

Applying the dominated convergence theorem (this can be used because the partial derivative of  $w$  satisfies  $\left| \frac{\partial w(x, y)}{\partial x} \right| \leq K$  when it exists, by Asm. 2.1), we can see that  $g$  is continuous on  $D_\lambda^c$ .

<sup>4</sup>We have seen already that  $D_\lambda$  is countable so in fact  $f'$  exists almost everywhere. This means that the values of  $f'$  where they exist will uniquely determine the element of  $L^2([0, 1])$  that is an eigenvalue-eigenfunction pair of  $\mathbb{M}$ .

Define  $f(x) = \int_0^x g(t)dt$  for all  $x \in [0, 1]$ . Then  $f'(x) = g(x)$  for  $x \in D_\lambda^\xi$ , and furthermore

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

Therefore,  $f - \frac{C}{\lambda}$  is an eigenfunction of  $\mathbb{L}$  as desired.

Finally, we check that  $0 \notin \sigma(\mathbb{M})$ . To see this, consider a function  $g$  such that  $\mathbb{M}g = 0$ . Then by the above calculation  $\mathbb{L}f = 0$ , where  $f(x) = \int_0^x g(t)dt - \frac{C}{\lambda}$ . Recalling from the paragraph following (2.4) that  $\lambda_1$  is a simple eigenvalue of  $\mathbb{L}$ , this implies  $f \in \text{Vect}(e)$  and so  $g = f' = 0$ . We conclude that  $\sigma(\mathbb{M}) = \sigma(\mathbb{L}) \setminus \{0\}$ .  $\square$

**Lemma 3.5.** *There exists  $f^* \in L^2([0, 1])$  such that  $\mathbb{M}f^* = \lambda_2 f^*$  and  $f^*(x) > 0$  for almost all  $x \in [0, 1]$ .*

*Proof.* Fix  $\alpha > 1 + \lambda_{\max}(\mathbb{M})$ . Define the operator  $\mathbb{M}_\alpha = \alpha\mathbb{I} - \mathbb{M}$  with  $\mathbb{I}$  the identity operator. It is obvious that  $\sigma(\mathbb{M}_\alpha) = \alpha - \sigma(\mathbb{M})$ , hence  $\lambda_{\max}(\mathbb{M}_\alpha) > 0$ .

We next claim that  $\mathbb{M}_\alpha$  sends the cone of nonnegative functions,

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

to itself. To see this, we examine the following expression for  $f \in K$  and  $x \in [0, 1]$ :

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

The first term  $(\alpha - d(x))f(x)$  is non-negative since  $\alpha > 1 \geq d(x)$ . For the second term, if  $y > x$  then  $\int_x^y f(z)dz \geq 0$  and  $\frac{\partial w(x, y)}{\partial x} \geq 0$  (because  $w$  is Robinsonian). Similarly, if  $y < x$ , both of these are negative and so their product is positive. Hence, in either case,  $\left( \int_x^y f(z)dz \right) \frac{\partial w(x, y)}{\partial x} \geq 0$  for all  $y$ . Therefore  $\mathbb{M}_\alpha f \in K$ , as required.

Consequently  $\mathbb{M}_\alpha$  is a positive, bounded and self-adjoint linear operator, with  $\mathbb{M}_\alpha K \subseteq K$ . The essential spectrum of  $\mathbb{M}_\alpha$  is  $\sigma_{\text{ess}}(\mathbb{M}_\alpha) = \alpha - rg(d)$  (because  $\sigma_{\text{ess}}(\mathbb{M}) = rg(d)$ ). Since  $\lambda_{\max}(\mathbb{M}_\alpha)$  is the largest eigenvalue of  $\mathbb{M}_\alpha$  and  $\sigma(\mathbb{M}_\alpha) = \alpha - \sigma(\mathbb{M})$ , additionally  $\sigma(\mathbb{M}) = \sigma(\mathbb{L}) \setminus \{0\}$  by Lemma 3.4 and so  $\lambda_{\max}(\mathbb{M}_\alpha) = \alpha - \lambda_2$ .

In [30] the authors showed a version of the Krein-Rutman theorem for non-compact operators, which applies here. In fact according to Theorem 1 from [30], there exists  $f^* \in K \setminus \{0\}$  such that  $\mathbb{M}_\alpha f^* = \lambda_{\max}(\mathbb{M}_\alpha) f^*$ , note also that the conditions of this theorem are satisfied since  $\lambda_{\max}(\mathbb{M}_\alpha) = \alpha - \lambda_2 > \alpha - \min_{x \in [0, 1]} d(x) = \sup \sigma_{\text{ess}}(\mathbb{M}_\alpha)$  according to Assumption 2.7. Note this is equivalent to  $\mathbb{M}f^* = \lambda_2 f^*$  as required.

We must now check that  $f^*(x) > 0$  for all  $x \in [0, 1]$ . Assume not; then there exists some  $a \in [0, 1]$  for which  $f^*(a) = 0$ . We then have  $(\mathbb{M}f^*)(a) = \lambda_2 f^*(a) = 0$ , which yields

$$\int_0^1 \left( \int_a^y f^*(z)dz \right) \frac{\partial w(a, y)}{\partial x} dy = 0. \quad (3.3)$$

Since the integrand is nonnegative, this implies that for almost all  $y \in [0, 1]$ ,

$$\left( \int_a^y f^*(z) dz \right) \frac{\partial w(a, y)}{\partial x} = 0. \tag{3.4}$$

By Assumption 2.2,  $\frac{\partial w(a, y)}{\partial x} \neq 0$  almost everywhere, so equation (3.4) implies  $\int_a^y f^*(z) dz = 0$  a.e. This means that  $f^* = 0$  in  $L^2([0, 1])$ , which contradicts the fact that  $f^*$  is an eigenfunction of  $\mathbb{M}$ . We conclude that no such  $a \in [0, 1]$  exists, so  $f^*(x) > 0$  for all  $x \in [0, 1]$  as desired.  $\square$

**Corollary 3.6.** *The Fiedler value  $\lambda_2 \notin rg(d)$ .*

*Proof.* Assume not, and let  $\lambda_2 = d(a)$  for some  $a$ . Let  $f^*$  be as in the statement of Lemma 3.5. Since  $\mathbb{M}f^*(a) = \lambda_2 f^*(a)$ , we get

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

Following the argument in the proof of Lemma 3.5 from equation (3.3), this implies  $f^* = 0$  in  $L^2([0, 1])$ . But we have already seen that  $f^*(x) > 0$  for all  $x \in [0, 1]$ , so this is a contradiction.  $\square$

We can now prove Theorem 3.1:

*Proof of Theorem 3.1.* By Corollary 3.6,  $\lambda_2 \notin rg(d)$ , and so by Lemma 3.3 all eigenfunctions of  $\mathbb{L}$  associated with  $\lambda_2$  are  $\mathcal{C}^1([0, 1])$ . Let  $f^*$  be as in the statement of Lemma 3.5, and define  $\phi(x) = \int_0^x f^*(t) dt$  for all  $x \in [0, 1]$ . By Lemma 3.4 we know that  $\mathbb{L}\phi = \lambda_2\phi$ , thus  $\phi \in \mathcal{C}^1([0, 1])$  is a Fiedler function with  $\phi'(x) = f^*(x) > 0$  for all  $x \in [0, 1]$ .

Finally, since  $\phi \in \mathcal{C}^1([0, 1])$  and  $\phi'(x) > 0$  for all  $x \in [0, 1]$ , we have  $\inf_{x \in [0, 1]} \phi'(x) > 0$ .  $\square$

Finally, we present the proof of Theorem 3.2:

*Proof of Theorem 3.2.* Let  $\phi$  as in the statement of Theorem 3.1. We prove Theorem 3.2 by contradiction. If the eigenspace is not one-dimensional, there exists some other Fiedler function  $\varphi \notin \text{Vect}(\phi)$ . Since  $\lambda_2 \notin rg(d)$  by Corollary 3.6, we have that  $\varphi \in \mathcal{C}^1([0, 1])$  as well. Let  $a$  be such that

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

Put  $\psi = \phi'(a)\varphi - \varphi'(a)\phi$ , so that  $\psi$  is another Fiedler function and satisfies  $\psi'(x) \geq 0$  for all  $x \in [0, 1]$  and  $\psi'(a) = 0$ . But this implies  $\mathbb{M}\psi'(a) = \lambda_2\psi'(a) = 0$ . Following the argument in the proof of Lemma 3.5 from equation (3.3), this implies that  $\psi = 0$  in  $L^2([0, 1])$ , which is a contradiction.  $\square$

#### 4. CONVERGENCE OF THE FIEDLER EIGENFUNCTION

We follow the conventions established at the start of Section 3, fixing a graphon  $w$  satisfying Assumptions 2.1 to 2.7 and defining the Laplacian  $\mathbb{L}$  as in equation (2.3). We will also use the notation for random graphs and their associated objects from Section 2.

In this section, we will check that the observed graph Laplacian converges to the limiting Laplacian  $\mathbb{L}$  in a strong enough sense to recover spectral data about the Fiedler eigenfunction.

#### 4.1. Convergence of laplacian operator

The results in this section are very similar to the results in the paper [14]. In this section we prove that two approximations to  $\mathbb{L}$  converge quickly. But first, we set some notation.

For an operator  $\mathbb{T}$  on  $L^2[0, 1]$ , define  $\|\mathbb{T}\| = \sup_{f \in L^2[0,1]: \|f\|=1} \int_0^1 \mathbb{T}f(x)^2 dx$ . Define  $d_{\text{MAX}} := \max_{x \in [0,1]} d(x)$  and similarly  $d_{\text{MAX}}^{(N)} := \max_{i=1}^N \left( \sum_{j=1}^N P_{ij}^{(N)} \right)$ . Let  $K$  be as in Assumption 2.1 and let  $\gamma > 1$ . Define the integers  $N_1(w, \tau, \varepsilon)$  and  $N'_1(w, \tau, \gamma)$  as follows:

$$N_1(w, \tau, \varepsilon) = \min \left\{ n \in \mathbb{N} : N \geq n \Rightarrow d_{\text{MAX}} \geq \frac{4 \log(4N/\varepsilon)}{9\rho_N N} + \frac{2K}{N} \right\}, \quad (4.1)$$

$$N'_1(w, \tau, \gamma) = \min \left\{ n \in \mathbb{N} : N \geq n \Rightarrow d_{\text{MAX}} \geq \frac{4 \log(4N/\varepsilon_\gamma(N))}{9\rho_N N} + \frac{2K}{N} \right\}. \quad (4.2)$$

Recall that  $\varepsilon_\gamma(N) = \exp(-\log^\gamma(N))$ . Note that both  $N_1$  and  $N'_1$  are well defined. In fact the term  $\frac{4 \log(2N/\varepsilon)}{9\rho_N N}$  goes to 0 as  $N \rightarrow \infty$  and we also have

$$\frac{4 \log(4N/\varepsilon_\gamma(N))}{9\rho_N N} = \frac{4(\log(4N) + \log^\gamma(N))}{9\rho_N N}$$

which goes to 0 as  $N \rightarrow \infty$ .

In order to prove the convergence of the Laplacian operator we need the following simple technical lemma:

**Lemma 4.1.** *Fix  $\varepsilon \in (0, 1)$ ,  $\gamma > 1$ . Put  $N_1 = N_1(w, \tau, \varepsilon)$  and  $N'_1 = N'_1(w, \tau, \gamma)$ . Then for all  $N \geq N_1(w, \tau, \varepsilon)$  we have*

$$d_{\text{MAX}}^{(N)} \geq \frac{4 \log(4N/\varepsilon)}{9\rho_N}.$$

Similarly, for all  $N \geq N'_1$  we have

$$d_{\text{MAX}}^{(N)} \geq \frac{4 \log(4N/\varepsilon_\gamma(N))}{9\rho_N}.$$

*Proof.* For  $N \geq N_1$ , let  $x \in \mathcal{I}_i^N$  for some  $i \in [N]$ . Then

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

where the inequality from line two to line three follows from Assumption 2.1. Therefore  $d_N(x) \geq d(x) - \frac{2K}{N}$ , so

$$\begin{aligned} \max_{x \in [0,1]} d_N(x) &\geq d_{\text{MAX}} - \frac{2K}{N} \\ &\geq \frac{4 \log(4N/\varepsilon)}{9\rho_N N}. \end{aligned}$$

On the other hand,

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

As required. The proof of the second inequality for  $N \geq N'_1(w, \tau, \gamma)$  is identical to the previous proof (replace  $\varepsilon$  by  $\varepsilon_\gamma(N)$ ).  $\square$

We have:

**Theorem 4.2** (Convergence of graphon Laplacian). *Let  $K$  be as in Assumption 2.1. Then*

$$\|\mathbb{L}_N - \mathbb{L}\| \leq \frac{4K}{N}. \quad (4.3)$$

Moreover, for all  $\varepsilon \in (0, 1)$  and all  $N \geq N_1(w, \tau, \varepsilon)$  as in equation (4.1),

$$\mathbb{P} \left( \left\| \rho_N^{-1} \widehat{\mathbb{L}}_N - \mathbb{L} \right\| \leq 4 \sqrt{\frac{\log(4N/\varepsilon)}{\rho_N N}} + \frac{4K}{N} \right) \geq 1 - \varepsilon. \quad (4.4)$$

Additionally, for all  $\gamma > 1$  the same inequality holds for all  $N \geq N'_1(w, \tau, \gamma)$  if we replace  $\varepsilon$  by  $\varepsilon_\gamma(N)$ , where  $N'_1$  is defined in (4.2).

*Proof of Theorem 4.2.* We prove the two inequalities (4.3) and (4.4) in order.

*Proof of Inequality (4.3).* Let  $f \in L^2([0, 1])$  such that  $\|f\| = 1$ . Then

$$\begin{aligned} \|\mathbb{L}_N f - \mathbb{L} f\|^2 &\leq \int_0^1 (\mathbb{L}_N f - \mathbb{L} f)^2(x) dx \\ &= \int_0^1 \left( \int_0^1 (f(x) - f(y))(w_N(x, y) - w(x, y)) dy \right)^2 dx \\ &\leq \int_0^1 \left( \int_0^1 (f(x) - f(y))^2 dy \right) \left( \int_0^1 (w_N(x, y) - w(x, y))^2 dy \right) dx \\ &= \sum_{i=1}^N \int_{\mathcal{I}_i^N} \left( \int_0^1 (f(x) - f(y))^2 dy \right) \left( \int_0^1 (w_N(x, y) - w(x, y))^2 dy \right) dx, \end{aligned} \quad (4.5)$$

where from line two to three we used Cauchy-Schwartz inequality. For fixed  $i \in [N]$  and  $x \in \mathcal{I}_i^N$ ,

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

Furthermore,

$$\begin{aligned} \int_0^1 (f(x) - f(y))^2 dy &\leq \int_0^1 2(f(x)^2 + f(y)^2) dy \\ &= 2f(x)^2 + 2. \end{aligned}$$

Applying the last two calculations with Inequality (4.5),

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

This immediately implies (4.3).  $\square$

Define  $D^{(N)}$  and  $\widehat{D}^{(N)}$  to be the row-sums of the matrices  $P^{(N)}$  and  $\widehat{P}^{(N)}$  respectively.

*Proof of Inequality (4.4).* From the triangle inequality and Inequality (4.3),

$$\left\| \rho_N^{-1} \widehat{\mathbb{L}}_N - \mathbb{L} \right\| \leq \left\| \rho_N^{-1} \widehat{\mathbb{L}}_N - \mathbb{L}_N \right\| + \left\| \mathbb{L}_N - \mathbb{L} \right\| \leq \left\| \rho_N^{-1} \widehat{\mathbb{L}}_N - \mathbb{L}_N \right\| + \frac{4K}{N}. \quad (4.6)$$

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

$$\begin{aligned} \left\| \rho_N^{-1} \widehat{\mathbb{L}}_N - \mathbb{L}_N \right\| &= \frac{1}{N} \lambda_{\max} \left( \rho_N^{-1} \widehat{L}^{(N)} - L^{(N)} \right) \\ &= \frac{\rho_N^{-1}}{N} \left\| \widehat{L}^{(N)} - \rho_N L^{(N)} \right\| \\ &\leq \frac{\rho_N^{-1}}{N} \left( \left\| \widehat{D}^{(N)} - \rho_N D^{(N)} \right\| + \left\| \widehat{P}^{(N)} - \rho_N P^{(N)} \right\| \right). \end{aligned} \quad (4.7)$$

Fix  $N \geq N_1(w, \tau, \varepsilon)$  as in (4.1), then the maximum expected degree of the random graph  $G^{(N)}$  is  $\rho_N d_{\text{MAX}}^{(N)} \geq \frac{4}{9} \log(4N/\varepsilon)$ . Applying Theorem 1 of [34],

$$\mathbb{P} \left( \left\| \widehat{P}^{(N)} - \rho_N P^{(N)} \right\| \leq \sqrt{4\rho_N d_{\text{MAX}}^{(N)} \log(4N/\varepsilon)} \right) \geq 1 - \frac{\varepsilon}{2}. \quad (4.8)$$

Now we give a similar bound for  $\left\| \widehat{D}^{(N)} - \rho_N D^{(N)} \right\|$ . Let  $\Delta^{(N)} = \widehat{D}^{(N)} - \rho_N D^{(N)}$  and  $X^{(N)} = \widehat{P}^{(N)} - \rho_N P^{(N)}$ .

We note that

$$\sum_j \text{Var}(X_{ij}) = \sum_j \rho_N P_{ij}^{(N)} (1 - \rho_N P_{ij}^{(N)}) \leq \rho_N \sum_j P_{ij}^{(N)} \leq \rho_N d_{\text{MAX}}^{(N)}. \quad (4.9)$$

Then by Bernstein's inequality, for each  $t > 0$

$$\mathbb{P} \left( |\Delta_i^{(N)}| > t \right) = \mathbb{P} \left( \left| \sum_{j=1}^N X_{ij}^{(N)} \right| > t \right) \leq 2 \exp \left( \frac{-\frac{1}{2}t^2}{\sum_j \text{Var}(X_{ij}) + \frac{1}{3}t} \right) \stackrel{\text{Ineq. (4.9)}}{\leq} 2 \exp \left( \frac{-\frac{1}{2}t^2}{\rho_N d_{\text{MAX}}^{(N)} + \frac{1}{3}t} \right).$$

Taking a union bound,

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

Choosing  $t = \sqrt{4\rho_N d_{\text{MAX}}^{(N)} \log(4N/\varepsilon)}$ ,

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

Recalling that  $\|\Delta^{(N)}\| = \|\widehat{D}^{(N)} - \rho_N D^{(N)}\|$  and  $d_{\text{MAX}}^{(N)} \leq N$ , this implies

$$\mathbb{P}\left(\|\widehat{D}^{(N)} - \rho_N D^{(N)}\| \leq \sqrt{4\rho_N N \log(4N/\varepsilon)}\right) \geq 1 - \frac{\varepsilon}{2}. \quad (4.10)$$

For  $\varepsilon \in (0, 1)$  we substitute (4.8) and (4.10) in (4.7) to get

$$\mathbb{P}\left(\left\|\rho_N^{-1}\widehat{\mathbb{L}}_N - \mathbb{L}_N\right\| \leq 4\sqrt{\frac{\log(4N/\varepsilon)}{\rho_N N}}\right) \geq 1 - \varepsilon. \quad (4.11)$$

Finally, applying inequality (4.6), we get that for all  $\varepsilon \in (0, 1)$  and all  $N \geq N_1(w, \tau, \varepsilon)$  in the sense of (4.1)

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

completing the proof. Again, replace in the  $\varepsilon$  by  $\varepsilon_\gamma(N)$  in the above and get the same result for all  $N \geq N'_1(w, \tau, \gamma)$  in the sense of (4.2).  $\square$

## 4.2. Convergence of the Fiedler function

In order to prove the convergence of the Fiedler function we use a result by Davis and Kahan [35] known as the  $\sin(\theta)$ -theorem for self-adjoint operators.

**Theorem 4.3** (Davis-Kahan). *Let  $\mathbb{A}$  and  $\widetilde{\mathbb{A}}$  be two self-adjoint operators with eigenvalues  $\lambda_1 < \lambda_2 < \lambda_3 < \dots$  and  $\tilde{\lambda}_1 < \tilde{\lambda}_2 < \tilde{\lambda}_3 < \dots$  respectively. Consider eigenfunctions  $f$  and  $\tilde{f}$  of  $\mathbb{A}$  and  $\widetilde{\mathbb{A}}$  respectively with eigenvalues  $\lambda_2$  and  $\tilde{\lambda}_2$  respectively, and normalized so that  $\|f\| = \|\tilde{f}\| = 1$  and  $\langle f, \tilde{f} \rangle \geq 0$ . If  $0 < \lambda_2 < \lambda_3$  (i.e.,  $\lambda_2$  is isolated), then*

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

Applying this to  $\mathbb{L}$  and  $\rho_N^{-1}\widehat{\mathbb{L}}_N$  immediately gives:

**Theorem 4.4.** *Let  $\phi$  and  $\hat{\phi}_N$  be the Fiedler functions of  $\mathbb{L}$  and  $\rho_N^{-1}\widehat{\mathbb{L}}_N$  respectively, normalized so that  $\|\phi\| = \|\hat{\phi}_N\| = 1$  and  $\langle \phi, \hat{\phi}_N \rangle \geq 0$ . Fix  $\eta < \frac{1-\tau}{3}$ , then for all  $\varepsilon \in (0, 1)$  there exists a positive constant  $C = C(w)$  and an integer  $N_0 = N_0(w, \tau, \eta, \varepsilon)$  so that*

$$\mathbb{P}\left(\left\|\hat{\phi}_N - \phi\right\| \leq C\sqrt{N^{-3\eta}}\right) \geq 1 - \varepsilon$$

for all  $N \geq N_0$ .

Moreover, for any  $\gamma > 1$  there exists a positive constant  $C' = C'(w)$  and an integer  $N'_0 = N'_0(w, \tau, \eta, \gamma)$  so that

$$\mathbb{P}\left(\left\|\hat{\phi}_N - \phi\right\| \leq C'\sqrt{N^{-3\eta}}\right) \geq 1 - \varepsilon_\gamma(N)$$

for all  $N \geq N'_0$ .

*Proof.* Define  $N_0 = N_0(w, \tau, \eta, \varepsilon)$  and  $N'_0 = N'_0(w, \tau, \eta, \gamma)$  by

$$N_0(w, \tau, \eta, \varepsilon) = \min \left\{ n \geq N_1(w, \tau, \varepsilon) : N \geq n \Rightarrow 4\sqrt{\frac{\log(4N/\varepsilon)}{\rho_N N}} + \frac{4K}{N} \leq \sqrt{N^{-3\eta}} \right\} \quad (4.12)$$

$$N'_0(w, \tau, \eta, \gamma) = \min \left\{ n \geq N'_1(w, \tau, \eta) : N \geq n \Rightarrow 4\sqrt{\frac{\log(4N/\varepsilon_\gamma(N))}{\rho_N N}} + \frac{4K}{N} \leq \sqrt{N^{-3\eta}} \right\} \quad (4.13)$$

where  $N_1$  and  $N'_1$  are defined in (4.1), (4.2) respectively. Note that both  $N_0$  and  $N'_0$  are well defined since the left hand sides of the inequalities inside  $N_0$  and  $N'_0$  respectively are on the order of  $\mathcal{O}(N^{-(1-\tau)/2} \log^{1/2}(N))$  and  $\mathcal{O}(N^{-(1-\tau)/2} \log^{\gamma/2}(N))$  respectively, which converges faster than the right hand sides of order  $\mathcal{O}(N^{-3\eta/2})$  (recall that  $\frac{3\eta}{2} < \frac{1-\tau}{2}$ ).

Take  $N \geq N_0$ , by Theorem 4.2 we know that with a probability of at least  $1 - \varepsilon$

$$\begin{aligned} \left\| \rho_N^{-1}\widehat{\mathbb{L}}_N - \mathbb{L} \right\| &\leq 4\sqrt{\frac{\log(4N/\varepsilon)}{\rho_N N}} + \frac{4K}{N} \\ &\leq \sqrt{N^{-3\eta}}. \end{aligned}$$

We apply the Davis-Kahan Theorem 4.3 to the Fiedler functions  $\phi$  and  $\hat{\phi}_N$  (note that the theorem applies here since under our assumptions  $\lambda_2$  is simple by Thm. 3.2). We have with probability at least  $1 - \varepsilon$

$$\left\| \hat{\phi}_N - \phi \right\| \leq \frac{2\sqrt{2} \left\| \rho_N^{-1}\widehat{\mathbb{L}}_N - \mathbb{L} \right\|}{\min\{\lambda_2, \lambda_3 - \lambda_2\}} \quad (4.14)$$

$$\leq \frac{2\sqrt{2}}{\min\{\lambda_2, \lambda_3 - \lambda_2\}} \sqrt{N^{-3\eta}} \quad (4.15)$$

as required. We get the same result for the case  $\varepsilon = \varepsilon_\gamma(N)$  and  $N \geq N'_0$ .  $\square$

## 5. ERROR BOUND FOR THE SPECTRAL SERIATION ALGORITHM

In this section, we prove Theorem 2.8. The method we use to bound the errors of the permutation are very similar to the method used in the paper [22]. We follow the notation used in the statement of Theorem 2.8, denote by  $\phi$  the Fiedler eigenfunction of  $\mathbb{L}$ , and denote by  $\hat{\phi}_N$  the Fiedler vector appearing in step 3 of Algorithm 1 with input  $G^{(N)}$ . By a small abuse of notation, for  $i \in [N]$  we define  $\hat{\phi}_i$  to be the  $i$ 'th element of  $\hat{\phi}_N$ . Similarly, for  $x \in [0, 1]$ , we define

$$\hat{\phi}_N(x) = \sum_{i=1}^N \hat{\phi}_i \mathbf{1}_{\mathcal{I}_i^N}(x),$$

where we recall  $\mathcal{I}_1^N = [0, 1/N]$  and  $\mathcal{I}_i^N = ((i-1)/N, i/N]$  for  $i = 2, \dots, N$ .

It is straightforward to check that the map  $x \mapsto \hat{\phi}_N$  is exactly the Fiedler function of  $\rho_N^{-1} \widehat{\mathbb{L}}_N$ . Note that we use the same notation  $\hat{\phi}_N$  to refer to both the Fiedler vector associated with finite graph  $G^{(N)}$  and the Fiedler function associated with the empirical graphon  $\rho_N^{-1} \widehat{\mathbb{L}}_N$ . This will not cause any ambiguity in the following argument, as it is always clear if the argument for  $\hat{\phi}_N$  is a generic value  $x \in [0, 1]$  or a generic value  $i \in \{1, 2, \dots, N\}$ . We repeat notation in this way because the two objects take exactly the same values in the same order, and this allows us to avoid some repetition of essentially identical calculations.

Let  $\sigma = \sigma_{\hat{\phi}_N}$  as in equation (1.5). Note that  $\sigma$  is the output of Algorithm 1 with input  $G^{(N)}$  and that we have  $\hat{\phi}_{\sigma^{-1}(1)} \leq \hat{\phi}_{\sigma^{-1}(2)} \leq \dots \leq \hat{\phi}_{\sigma^{-1}(N)}$ . Since  $\phi$  is monotone (by Thm. 3.1) and  $\hat{\phi}_N$  is close to  $\phi$  or  $-\phi$  (by Thm. 4.4), we expect the permutation  $\sigma$  to be close to the identity permutation (or its reverse).

For any permutation  $\pi$  in the symmetric group  $\mathcal{S}_N$ , define the Kendall tau distance to be

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

In [36], it was shown that, for any  $\pi \in \mathcal{S}_N$ , we have  $D_N(\pi) \leq \|\pi - id_N\|_1 \leq 2D_N(\pi)$ . Thus, when  $N \rightarrow \infty$ , these distances are of the same order.

We are now ready to prove Theorem 2.8:

*Proof of Theorem 2.8.* Fix any  $0 < \eta < \frac{1-\tau}{3}$ . We assume without loss of generality that  $\|\hat{\phi}_N - \phi\| \leq \|\hat{\phi}_N + \phi\|$  and that  $\phi$  is monotone increasing (if not, replace  $\phi$  with  $-\phi$  and/or  $\hat{\phi}_N$  by  $-\hat{\phi}_N$  in the following). Define

$$\begin{aligned} E_N &:= \{(i, j) \in [N]^2 : i < j \text{ and } \sigma(j) < \sigma(i)\} \\ E_N^\eta &:= \{(i, j) \in [N]^2 : i + N^{1-\eta} < j \text{ and } \sigma(j) < \sigma(i)\}, \end{aligned}$$

We then have

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

Since  $\phi(x + \frac{j-i}{N}) > \phi(x)$  and  $\hat{\phi}_j \leq \hat{\phi}_i$  for  $(i, j) \in E_N$ , we also have

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

Combining the previous two estimates,

$$\begin{aligned} 2 \left\| \hat{\phi}_N - \phi \right\|^2 &\geq \frac{1}{2N} \sum_{(i,j) \in E_N^\eta} \int_{\mathcal{I}_i^N} \left( \phi(x + \frac{j-i}{N}) - \phi(x) \right)^2 dx \\ &\stackrel{\text{Theorem 3.1}}{\geq} \frac{L^2}{2N^3} \sum_{(i,j) \in E_N^\eta} \int_{\mathcal{I}_i^N} (j-i)^2 dx \\ &\geq \frac{L^2}{2N^{1+2\eta}} \sum_{(i,j) \in E_N^\eta} \int_{\mathcal{I}_i^N} dx \\ &= \frac{L^2}{2N^{2+2\eta}} |E_N^\eta|, \end{aligned}$$

where the inequality from line two to three follows from the fact that  $j > i + N^{1-\eta}$  for  $(i, j) \in E_N^\eta$ . Consequently,

$$|E_N^\eta| \leq \frac{4}{L^2} N^{2+2\eta} \left\| \hat{\phi}_N - \phi \right\|^2.$$

On the other hand,

$$E_N \setminus E_N^\eta = \{(i, j) \in [N]^2 : 0 < j - i \leq N^{1-\eta} \text{ and } \sigma(j) < \sigma(i)\},$$

thus  $|E_N \setminus E_N^\eta| \leq N^{2-\eta}$ . By Theorem 4.4 there exists a positive constant  $C = C(w)$  so that for all  $N \geq N_0(w, \tau, \eta, \varepsilon)$  we have with probability at least  $1 - \varepsilon$

$$\left\| \hat{\phi}_N - \phi \right\|^2 \leq C^2 N^{-3\eta}.$$

Recall that  $D_N(\sigma) = |E_N|$  and  $\|\sigma - id_N\|_1 \leq \frac{1}{2} D_N(\sigma)$ , therefore with probability at least  $1 - \varepsilon$

$$\|\sigma - id_N\|_1 \leq 2(|E_N \setminus E_N^\eta| + |E_N^\eta|) \tag{5.1}$$

$$\leq 2N^{2-\eta} + \frac{4}{L^2} N^{2+2\eta} \left\| \hat{\phi}_N - \phi \right\|^2 \tag{5.2}$$

$$\leq \left(1 + \frac{4C^2}{L^2}\right) N^{2-\eta} \tag{5.3}$$

as required. The same inequality holds for  $\varepsilon = \varepsilon_\gamma(N)$  for all  $N \geq N'_0(w, \tau, \eta, \gamma)$ .  $\square$

## 6. POSTPROCESSING AND TIGHTER ERROR BOUNDS

The main goal of this section is to prove Theorem 2.13. We introduce a post-processing algorithm of the estimate returned by Algorithm 1. This new algorithm provides an estimate that is more robust to noise, with a nearly optimal convergence rate.

### 6.1. Notation

We set notation that will be used throughout Section 6. Let  $G = (V, E)$  be a graph of size  $N$ . Denote by  $V := \{v_1, v_2, \dots, v_N\}$  the vertices of  $G$ , with  $v_i = i/N$ . For any  $S \subset V$ , denote the *induced subgraph* on  $S$  by  $G|_S$ . Define the (restricted) *neighbourhood* of a vertex  $v \in V$  by  $N_S(v) = \{u \in S : (u, v) \in E\}$ . Denote by  $\sigma_S : S \rightarrow [|S|]$  the function that sends  $s \in S$  to  $|\{s' \in S : s' \leq s\}|$ ; we think of this ordering as being essentially the “identity” permutation on  $S$ .

Algorithm 3 involves partitioning up the vertices of  $G$ , estimating various quantities on parts of the partition, and then combining them. We say that a partition  $(S_1, S_2, S_3)$  of  $V$  is “good” if  $n \leq |S_3| \leq |S_2| \leq |S_1| \leq n + 1$ , where  $n = \lfloor \frac{N}{3} \rfloor$ .

Let  $(S_1, S_2, S_3)$  be a “good” partition of  $V$  and let  $\sigma_j : S_j \rightarrow [|S_j|]$  be orderings of each part of the partition. We define the “merged ordering” of  $(\sigma_1, \sigma_2, \sigma_3)$  as an ordering  $\sigma : [N] \rightarrow [N]$  defined by:

$$\sigma^{-1}(3k + j) = \sigma_j^{-1}(k + 1), \tag{6.1}$$

for all  $k \in \{0, 1, \dots, n\}$  and  $j \in \{1, 2, 3\}$  such that  $3k + j \leq N$ . Note that equation (6.1) does define a permutation, since for all  $i \in [N]$  there is exactly one solution to the equation  $i = 3k + j$  with  $j \in \{1, 2, 3\}$  and  $k \in \{0, 1, \dots, n\}$ . Note that the factor  $N$  in the definition is coming from the fact that the elements of  $V$  are scaled by  $v_i = i/N$ .

Finally, for two orderings  $\sigma_1, \sigma_2$  on sets  $S_1, S_2$ , say that  $\sigma_1, \sigma_2$  are *aligned* if

$$\max_{i \in S_1 \cap S_2} \left| \frac{\sigma_1(i)}{|S_1|} - \frac{\sigma_2(i)}{|S_2|} \right| \leq \max_{i \in S_1 \cap S_2} \left| \frac{\text{rev}(\sigma_1)(i)}{|S_1|} - \frac{\sigma_2(i)}{|S_2|} \right|, \tag{6.2}$$

and that  $\sigma_1, \sigma_2$  *not aligned* otherwise. If  $S_1 \cap S_2 = \emptyset$ , say that they are aligned. Moreover, if the left-hand side of (6.2) is less than 0.001 and the right-hand side is more than 0.999, say that  $\sigma_1, \sigma_2$  are *closely aligned*. Recall that  $\text{rev}(\sigma)$  denotes the reversal of any ordering  $\sigma$  defined in equation (1.2).

### 6.2. Algorithms and informal analysis

We are now ready to present our main algorithms.

We give an informal summary of what this algorithm is doing and why we expect it to work well, eliding many details. While the algorithms themselves are fairly long, the key simple idea is explained in Figure 1: the functions  $\Phi_R, \Phi_L$  are each monotone over large parts of  $[0, 1]$ , so we can easily compare the latent positions  $U_i, U_j$  of vertices  $i, j$  if we can (i) check which of  $\Phi_R, \Phi_L$  to use in comparing them, and (ii) find a good approximation of  $\Phi_R(U_i), \Phi_R(U_j), \Phi_L(U_i)$  and  $\Phi_L(U_j)$ . Most of the length of the algorithm comes from case analysis (that is: determining which function to use when comparing a particular pair of vertices) and manufacturing a certain amount of independence between estimators used in different stages.

The algorithm accepts as input a partition of  $V$  into two subsets  $S, T$ . Steps 1-2 get a good estimate of the right-most and left-most vertices in  $T$ , denoted  $R$  and  $L$  respectively. Using these estimated sets, steps 3-6 try to compare vertices  $u, v \in S$  by comparing the number of neighbours these vertices have in  $R$  and  $L$ . The Robinsonian property suggests the following heuristic: if  $u < v$ , then  $u$  should have more neighbours in  $L$  while  $v$  should have more neighbours in  $R$ . If this heuristic were exactly correct, we would use only the first of four cases in the function  $\hat{F}$  from equation (6.9). However, the heuristic isn’t quite correct: it can fail if (i) both of  $u, v$  are very close to either 0 or 1 (in which case many of the “right-most” elements of  $R$  are not actually to the right of  $u, v$  or many of the “left-most” elements of  $L$  are not actually to the left of  $u, v$ ) or (ii) if  $|u - v|$  is too small (in which case sampling error may dominate).

The four cases in equation (6.9) deal with the first of these problems by checking if  $u, v$  are very close to either 0 or 1, which can be detected by comparing the values of the functions  $\hat{\Psi}_R, \hat{\Psi}_L$  to the empirical quantiles  $c_R, c_L$  computed in equations (6.7) and (6.8). This ensures that we compare  $u, v$  by counting either right- or left-most

---

**Algorithm 2:** Initial Post-processing the output of spectral seriation.

---

**Input :** A graph  $G = (V, E)$  of size  $N$ , a pair of disjoint sets  $(T, S) \subset V^2$  and parameters  $0 < \alpha < \beta < 1/2$ .

- 1 Let  $\hat{\sigma}_T$  be the result of running Algorithm 1 on  $G|_T$ .
- 2 Given  $\hat{\sigma}_T$ , estimate the set of right-most and left-most vertices in  $T$  by

$$R = R_{\hat{\sigma}_T}(\alpha) := \{u \in T : \hat{\sigma}_T(u) \geq (1 - \alpha) |T|\}, \quad (6.3)$$

$$L = L_{\hat{\sigma}_T}(\alpha) := \{u \in T : \hat{\sigma}_T(u) \leq \alpha |T|\}. \quad (6.4)$$

- 3 For each  $v \in S$ , estimate the neighbourhood-size functions

$$\hat{\Psi}_R(v) := \frac{1}{|T|} |N_R(v)| = \frac{1}{|T|} \sum_{u \in R} \mathbf{1}_{(u,v) \in E}, \quad (6.5)$$

$$\hat{\Psi}_L(v) := \frac{1}{|T|} |N_L(v)| = \frac{1}{|T|} \sum_{u \in L} \mathbf{1}_{(u,v) \in E}. \quad (6.6)$$

- 4 Compute the following empirical quantiles of  $\hat{\Psi}_L, \hat{\Psi}_R$ :

$$c_R := \inf \left\{ c \geq 0 : \left| \left\{ v \in S : \hat{\Psi}_R(v) \leq c \right\} \right| \geq (1 - \beta) |S| \right\}, \quad (6.7)$$

$$c_L := \inf \left\{ c \geq 0 : \left| \left\{ v \in S : \hat{\Psi}_L(v) \leq c \right\} \right| \geq (1 - \beta) |S| \right\}. \quad (6.8)$$

- 5 Compute the antisymmetric function  $\hat{F} : S^2 \mapsto \{-1, 0, 1\}$  by the following rule for  $u < v$ :

$$\hat{F}(u, v) = \left\{ \begin{array}{ll} 1 - 2 \cdot \mathbf{1}_{\hat{\Psi}_R(u) > \hat{\Psi}_R(v)} & \text{if } \hat{\Psi}_R(u), \hat{\Psi}_R(v) < c_R; \text{ otherwise,} \\ 1 - 2 \cdot \mathbf{1}_{\hat{\Psi}_L(u) < \hat{\Psi}_L(v)} & \text{if } \hat{\Psi}_L(u), \hat{\Psi}_L(v) < c_L; \text{ otherwise,} \\ 1 & \text{if } \hat{\Psi}_R(u) < c_R; \text{ otherwise,} \\ -1 & \end{array} \right\}. \quad (6.9)$$

- 6 For each  $v \in S$  compute

$$\hat{f}(v) = \sum_{u \in S} \hat{F}(u, v). \quad (6.10)$$

- 7 Compute the ordering  $\hat{\sigma}_S = \hat{\sigma}_{\hat{f}}$  on  $S$  according to equation (1.5), breaking ties arbitrarily.

**Output:** The permutation  $\hat{\sigma}_S$ .

---

neighbours as appropriate. To give a back-of-the-envelope analysis of the sampling error that can cause the second problem, we note that there are  $\Theta(|T|)$  possible edges from  $u, v$  to  $R$  or  $L$ , and the presence or absence of these edges are independent of the random variables used to compute  $R$  and  $L$  (the purpose of partitioning  $V$  is exactly to obtain this independence). This suggests that, as long as the error in the estimates  $R, L$  of  $R_{id_N}$  and  $L_{id_N}$  are “negligible”, the typical error in the estimates  $\hat{\Psi}_R, \hat{\Psi}_L$  should be  $\mathcal{O}(\sqrt{N})$ . In particular, this is the step that lets us clean rough estimates of sets  $R_{id_N}, L_{id_N}$  into much better estimates of the averages  $\hat{\Psi}_{R_{id_N}}, \hat{\Psi}_{L_{id_N}}$ . Although this may not be obvious at first glance, the consistency result in Theorem 2.8 is strong enough to guarantee that the error in the estimates of  $R_{id_N}, L_{id_N}$  are “negligible” in the sense that this argument requires.

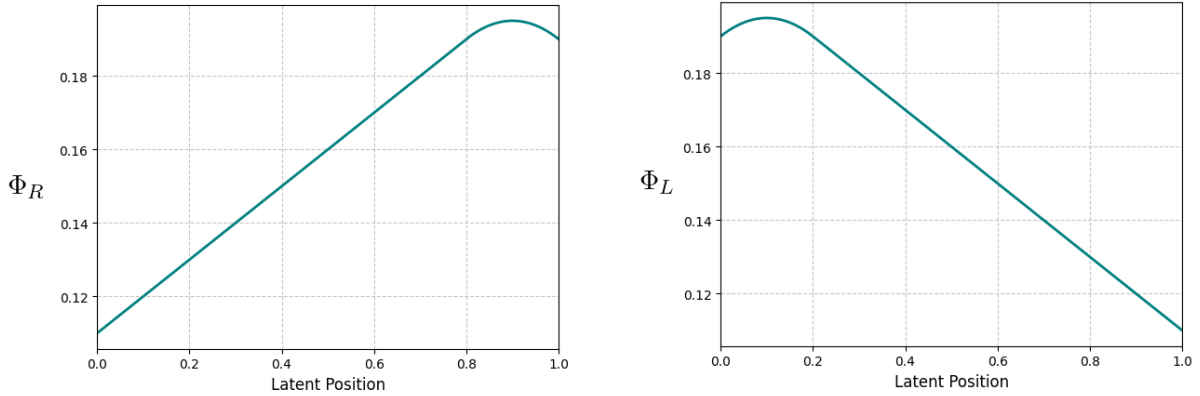


FIGURE 1. Plots of  $\Phi_R, \Phi_L$  for  $w(x, y) = 1 - 0.5|x - y|$  and  $\alpha = 0.2$ . Note that the test functions  $\Phi_R, \Phi_L$  are both monotone over intervals  $I_R, I_L$  that together cover all of  $[0, 1]$ . Thus, we can use  $\Phi_R$  to order vertices  $i, j$  with  $U_i, U_j \in I_R$ , as long as we can both (i) estimate  $\Phi_R$  and (ii) determine that  $U_i, U_j$  are in the  $I_R$ . The same observation suffices for ordering vertices  $U_i, U_j \in I_R$ , and of course if  $U_i \in I_L \setminus I_R$  but  $U_j \notin I_L$  we can use that fact to order  $U_i, U_j$ .

This calculations suggests that  $\hat{\sigma}_S$  should have a sup-norm error that is not too much larger than  $\mathcal{O}(\sqrt{N})$ , which is the case.

We now give our overall post-processing algorithm:

---

**Algorithm 3:** Post-processing the output of spectral seriation.

---

**Input :** A graph  $G = (V, E)$  of size  $N$ . Parameters  $0 < \alpha < \beta < 1/2$ .

- 1 Choose  $(S_1, S_2, S_3)$  a “good” partition of  $V$ , uniformly at random; and independently choose another “good” partition  $(S'_1, S'_2, S'_3)$ , uniformly at random.
- 2 Use Algorithm 2 on pairs of subsets  $(S_3, S_1)$ ,  $(S_1, S_2)$  and  $(S_2, S_3)$  with parameters  $\alpha, \beta$  to obtain estimated orders  $\hat{\sigma}_1, \hat{\sigma}_2, \hat{\sigma}_3$  on  $S_1, S_2, S_3$  respectively. Also use Algorithm 2 on  $(S'_1, S'_2)$  with parameters  $\alpha, \beta$  to obtain estimated order  $\hat{\sigma}_{\text{ref}}$  on  $S'_2$ .
- 3 For  $j \in \{1, 2, 3\}$ , replace  $\hat{\sigma}_j$  by  $\text{rev}(\hat{\sigma}_j)$  if it is not aligned with  $\hat{\sigma}_{\text{ref}}$  in the sense of equation (6.2).
- 4 Compute  $\hat{\sigma}$  the merged ordering of  $(\hat{\sigma}_1, \hat{\sigma}_2, \hat{\sigma}_3)$  according to equation (6.1).

**Output:** The permutation  $\hat{\sigma}$ .

---

Algorithm 3 largely consists of many calls to Algorithm 2, and so most of the algorithm seems straightforward. The one possibly-mysterious part is step 3. To explain the need for this step, recall that it is impossible to distinguish between a vertex ordering  $\sigma$  and its reverse  $\text{rev}(\sigma)$  by looking only at the edges of a graph. Up to this point in the paper, this has not posed a problem, as we always looked at an entire graph all at once. However, in Algorithm 3, we use Algorithm 2 to repeatedly order different subsets of the vertices. As a result, it is likely that some of the returned orderings  $\hat{\sigma}_j : S_j \rightarrow [|S_j|]$  will be aligned with  $\sigma_{S_j}$  (in the sense of Eq. (6.2)) while others will be aligned with  $\text{rev}(\sigma_{S_j})$ . Fortunately, this problem is not difficult to solve: each ordering will either be closely aligned with  $\sigma_{S_j}$  or its reverse, and so a single reference ordering  $\hat{\sigma}_{\text{ref}}$  can be used to simultaneously align *all* other orderings, as long as their supports have non-empty intersection.

### 6.3. Analysis of algorithm 2

In this subsection, we prove the main bounds on the objects appearing in Algorithm 3. We start by presenting the main theorem that provides a bound for the error of the estimated ordering  $\hat{\sigma}_S$  returned by Algorithm 2. Then the rest of this subsection will be dedicated to proving this theorem.

**Theorem 6.1.** *Let  $G = (V, E)$  be a Robinsonian graph of size  $N$  sampled according to a graphon  $w$  as in Theorem 2.13. Fix  $s, t$  satisfying  $n \leq s, t \leq n + 1$  and let  $S, T \subset V$  be chosen uniformly at random from all disjoint subsets of size  $|S| = s, |T| = t$ . Consider  $\hat{\sigma}_S : S \rightarrow [|S|]$  to be the output of Algorithm 2 with input  $G$ , the pair of random sets  $(T, S)$  and parameters  $0 < \alpha < \beta < 1/2$  as in Assumption 2.10. For all  $\gamma > 1$  and  $\epsilon > 0$ , there exists a positive constant  $\delta = \delta(w, \gamma, \epsilon) > 0$  and an integer  $N_0 = N_0(w, \gamma, \epsilon)$  so that*

$$\mathbb{P}\left(\|\hat{\sigma}_S - \sigma_S\|_\infty \leq \epsilon \sqrt{N \log^\gamma(N)}\right) \geq 1 - \epsilon_\gamma^\delta(N)$$

for all  $N \geq N_0$ .

Throughout this subsection we use notation as in Theorem 6.1, let  $0 < \alpha < \beta < 1/2$  as in Assumption 2.10, and set  $\mu = (\alpha + \beta)/2$ .

We fix a constant  $\gamma > 1$  for the remainder of the section, and say that a sequence of events  $\{\mathcal{E}_N\}_{N \in \mathbb{N}}$  holds with *extreme probability* if there exist a constant  $\delta > 0$  and an integer  $N_0$ , such that

$$\mathbb{P}(\mathcal{E}_N) \geq 1 - \epsilon_\gamma^\delta(N) \tag{6.11}$$

for all  $N \geq N_0$ . We abbreviate this *w.e.p.* Note that for any  $C > 0$ , the intersection of  $\mathcal{O}(N^C)$  events that occur *w.e.p.* for some  $\gamma > 1$  also occur *w.e.p.* with the same  $\gamma$ . We define the sequence  $\Delta_N = \sqrt{\frac{\log^\gamma(N)}{N}}$  for  $N \geq 1$ ; this is essentially the scale of the most important errors that occur in the algorithm.

The first step in the proof is checking that the discrete functions  $\hat{\Psi}_R, \hat{\Psi}_L$  are good approximations of the continuous functions  $\Psi_R, \Psi_L$  respectively. We recall some facts about these functions. For  $v \in V$ , recall

$$\Psi_R(v) = \int_{1-\alpha}^1 w(v, u) du.$$

Denote by  $\mathcal{F}_T$  the  $\sigma$ -algebra generated by the choice of  $T$  and the graph  $G^{(N)}|_T$ . Since  $\hat{\Psi}_R(v)$  and  $\hat{\Psi}_L(v)$  are sums of Bernoulli random variables sampled according to equation (2.1) then

$$\mathbb{E}[\hat{\Psi}_R(v)|\mathcal{F}_T] = \frac{1}{|T|} \sum_{u \in R} w(u, v), \text{ and } \mathbb{E}[\hat{\Psi}_L(v)|\mathcal{F}_T] = \frac{1}{|T|} \sum_{u \in L} w(u, v);$$

in particular  $R, L$  are both  $\mathcal{F}_T$ -measurable, while the indicator function of an edge  $(u, v)$  with  $u \in L \cup R, v \in S$  is *independent* of  $\mathcal{F}_T$  and has mean  $w(u, v)$ . This implies that for any fixed  $\epsilon > 0$  the following holds *w.e.p.*

$$\max_{v \in S} \left| \hat{\Psi}_R(v) - \frac{1}{|T|} \sum_{u \in R} w(u, v) \right| \leq \epsilon \Delta_N. \tag{6.12}$$

The following Lemma shows that the graph-dependent functions  $\hat{\Psi}_R, \hat{\Psi}_L$  approximate the graphon-dependent functions  $\Psi_R, \Psi_L$

**Lemma 6.2.** *For any  $\epsilon > 0$ , the inequalities*

$$\max_{v \in S} \left| \hat{\Psi}_R(v) - \Psi_R(v) \right| \leq \epsilon \Delta_N, \text{ and } \max_{v \in S} \left| \hat{\Psi}_L(v) - \Psi_L(v) \right| \leq \epsilon \Delta_N$$

hold *w.e.p.*

Similarly, the constants  $c_R$  and  $c_L$  are intended to estimate  $\Psi(1 - \beta)$  and  $\Psi(\beta)$  respectively.

**Lemma 6.3.** *For any  $\epsilon > 0$  the following inequalities*

$$|c_R - \Psi_R(1 - \beta)| \leq \epsilon \Delta_N, \quad \text{and} \quad |c_L - \Psi_L(\beta)| \leq \epsilon \Delta_N$$

hold w.e.p.

*Proof of Lemma 6.2.* We will present the proof only for  $\hat{\Psi}_R$ , since the proof for  $\hat{\Psi}_L$  is identical. Denote by  $T = \{t_{(1)}, \dots, t_{(t)}\}$  the elements of  $T$  in order, so that  $t_{(1)} < t_{(2)} < \dots < t_{(t)}$ . Note that  $\sigma_T(t_{(k)}) = k$  for all  $k$ . For  $v \in S$ , we have

$$\mathbb{E}[\hat{\Psi}_R(v)|\mathcal{F}_T] = \frac{1}{t} \sum_{k=\lceil(1-\beta)t\rceil}^t w(t_{(k)}, v).$$

Now define

$$\Sigma := \frac{1}{N} \sum_{k=\lceil(1-\beta)N\rceil}^N w(v_k, v),$$

and, for each  $j = 0, 1, 2$ , define

$$\Sigma_j := \frac{1}{t} \sum_{k=\lceil(1-\beta)t\rceil}^t w(v_{3k+j}, v).$$

By the triangle inequality,

$$\left| \mathbb{E}[\hat{\Psi}_R(v)|\mathcal{F}_T] - \Psi_R(v) \right| \leq |\Psi_R(v) - \Sigma| + \frac{1}{3} \sum_{j=0}^2 \left| \mathbb{E}[\hat{\Psi}_R(v)|\mathcal{F}_T] - \Sigma_j \right| + \frac{1}{3} |\Sigma_0 + \Sigma_1 + \Sigma_2 - 3\Sigma|. \quad (6.13)$$

We now bound each term of the right hand side in inequality (6.13). Recall the constant  $K$  from Assumption 2.1. We have

$$|\Psi_R(v) - \Sigma| \leq \frac{\beta K}{N}, \quad (6.14)$$

by essentially the same argument as the proof of the first inequality in the proof of Lemma 4.1. Next,

$$\begin{aligned} |\Sigma_0 + \Sigma_1 + \Sigma_2 - 3\Sigma| &\leq \left| \Sigma_0 + \Sigma_1 + \Sigma_2 - \frac{N}{t} \Sigma \right| + \frac{|N - 3t|}{t} \Sigma \\ &= \frac{1}{t} \left| \underbrace{\sum_{k=\lceil 3(1-\beta)t \rceil}^N w(v_k, v) - \sum_{k=\lceil (1-\beta)N \rceil}^N w(v_k, v)}_{\leq (1-\beta)|N-3t|} \right| + \underbrace{\Sigma}_{\leq \beta} \frac{|N - 3t|}{t} \\ &\leq \frac{|N - 3t|}{t}, \end{aligned}$$

where the bound in the first underbrace comes from the facts that (i) at most  $(1 - \beta)|N - 3t|$  terms appear in one sum but not the other and (ii) each summands  $w(v_k, v)$  is bounded by 1, and the bound in the second

underbrace comes again from the fact that the summands of  $\Sigma$  are bounded by 1. Since  $3t - 3 \leq N \leq 3t + 2$  by assumption, we conclude

$$|\Sigma_0 + \Sigma_1 + \Sigma_2 - 3\Sigma| \leq \frac{12}{N}. \quad (6.15)$$

Finally, fixing  $0 \leq j \leq 2$ , we calculate

$$\begin{aligned} \left| \mathbb{E}[\hat{\Psi}_R(v)|\mathcal{F}_T] - \Sigma_j \right| &\leq \frac{1}{t} \sum_{k \geq (1-\beta)t} |w(t_{(k)}, v) - w(v_{3k+j}, v)| \\ &\leq \frac{K}{t} \sum_{k \geq (1-\beta)t} |t_{(k)} - v_{3k+j}|. \end{aligned}$$

According to Lemma A.1,

$$\max_{k \in T} |t_{(k)} - \frac{k}{t}| \leq \frac{\epsilon}{2\beta K} \Delta_N \quad w.e.p.$$

Additionally, it is easy to check that  $|\frac{k}{t} - v_{3k+j}| \leq 8/N$ . Therefore we have with extreme probability

$$|t_{(k)} - v_{3k+j}| \leq \frac{\epsilon}{2\beta K} \Delta_N + \frac{8}{N}.$$

We conclude

$$\left| \mathbb{E}[\hat{\Psi}_R(v)|\mathcal{F}_T] - \Sigma_j \right| \leq \frac{\epsilon}{2} \Delta_N + \frac{8\beta K}{N} \quad w.e.p. \quad (6.16)$$

We substitute Inequalities (6.14), (6.15) and (6.16) in Inequality (6.13) to get

$$\left| \mathbb{E}[\hat{\Psi}_R(v)|\mathcal{F}_T] - \Psi_R(v) \right| \leq \frac{\epsilon}{2} \Delta_N + \frac{9\beta K + 12}{N} \leq \epsilon \Delta_N \quad w.e.p.$$

Applying the concentration bound (6.12) completes the proof.  $\square$

Define

$$M = \sup_{x \in [0,1]} \Psi'_R(x) > 0; \quad (6.17)$$

this exists by Assumption 2.2.

*Proof of Lemma 6.3.* We give the proof only for  $\Psi_R$ , as the case  $\Psi_L$  is identical. Recall that by definition of  $c_R$

$$\left| \left\{ v \in S : \hat{\Psi}_R(v) \leq c \right\} \right| \geq (1-\beta)s \Leftrightarrow c \geq c_R$$

for all  $c > 0$ . For  $v \in V$ , we have

$$\{|v - (1-\beta)| \leq \frac{\epsilon \Delta_N}{2M}\} \Rightarrow \{|\Psi_R(v) - \Psi_R(1-\beta)| \leq \frac{\epsilon \Delta_N}{2}\}.$$

By Lemma 6.2, we have

$$|\hat{\Psi}_R(v) - \Psi_R(v)| \leq \frac{\epsilon \Delta_N}{2} \quad w.e.p.$$

Combining the last two bounds,

$$|v - (1 - \beta)| \leq \frac{\epsilon \Delta_N}{2M} \Rightarrow |\hat{\Psi}_R(v) - \Psi_R(1 - \beta)| \leq \epsilon \Delta_N \quad w.e.p.$$

Let  $\{s_{(k)}\}_{k=1}^s$  be the elements of  $S$  in increasing order and let  $p = \lceil (1 - \beta)s \rceil$ . By Lemma A.1,

$$|s_{(p)} - (1 - \beta)| \leq |s_{(p)} - \frac{p}{s}| + |\frac{p}{s} - (1 - \beta)| \leq \frac{\epsilon \Delta_N}{2M} \quad w.e.p.$$

By part 1 of Assumption 2.10,

$$\hat{\Psi}_R(s_{(k)}) \leq \Psi_R(1 - \beta) + \epsilon \Delta_N, \quad w.e.p.,$$

for all  $1 \leq k \leq p$  and

$$\Psi_R(1 - \beta) - \epsilon \Delta_N \leq \hat{\Psi}_R(s_{(k)}) \quad w.e.p.,$$

for all  $p \leq k \leq s$ . Therefore

$$\left| \left\{ v \in S : \hat{\Psi}_R(v) \leq \Psi_R(1 - \beta) + \epsilon \Delta_N \right\} \right| \geq p \geq (1 - \beta)s, \quad w.e.p.,$$

which implies that

$$c_R \leq \Psi_R(1 - \beta) + \epsilon \Delta_N \quad w.e.p. \quad (6.18)$$

On the other hand,

$$\left| \left\{ v \in S : \hat{\Psi}_R(v) \geq \Psi_R(1 - \beta) - \epsilon \Delta_N \right\} \right| \geq s - p + 1 > \beta s \quad w.e.p.,$$

this is equivalent to

$$\left| \left\{ v \in S : \hat{\Psi}_R(v) \leq \Psi_R(1 - \beta) - \epsilon \Delta_N \right\} \right| < (1 - \beta)s \quad w.e.p.,$$

which implies that

$$c_R > \Psi_R(1 - \beta) - \epsilon \Delta_N \quad w.e.p. \quad (6.19)$$

Combining Inequalities (6.18) and (6.19), we deduce that

$$|c_R - \Psi(1 - \beta)| \leq \epsilon \Delta_N \quad w.e.p.$$

The proof for  $c_L$  is identical to the proof of  $c_R$ . □

The following proposition states that if we have two vertices  $u, v \in S$  such that (i)  $u$  and  $v$  are not *very close* to each other, and (ii) both of them belong to one of the intervals  $[0, 1 - \mu]$  or  $[\mu, 1]$ . Then these vertices can be ordered accurately using the graph-dependent functions  $\hat{\Psi}_R$  and  $\hat{\Psi}_L$  *w.e.p.*

**Proposition 6.4.** *Let  $\epsilon > 0$ . With extreme probability, the events*

$$\begin{aligned} \{0 \leq u < v \leq 1 - \mu, \text{ and } v > u + \epsilon \Delta_N\} &\Rightarrow \{\hat{\Psi}_R(v) \geq \hat{\Psi}_R(u)\}, \\ \{\mu \leq u < v \leq 1, \text{ and } v > u + \epsilon \Delta_N\} &\Rightarrow \{\hat{\Psi}_L(v) \leq \hat{\Psi}_L(u)\} \end{aligned}$$

hold for all pairs  $(u, v) \in S^2$ .

*Proof of Proposition 6.4.* Recall that  $\hat{\sigma}_T$  from Algorithm 2 is the ordering obtained from Algorithm 1 with input  $G|_T$ . By Theorem 2.8, we know that for any  $\eta \in (0, 1/3)$  there exists a constant  $C > 0$  such that

$$\|\hat{\sigma}_T - \sigma_T\|_1 \leq C t^{2-\eta} \quad (6.20)$$

holds *w.e.p.* Note that the difference between  $\sigma_T$  and its reverse is on the order of  $t^2$ . This means that, when Inequality (6.20) holds and  $N$  is sufficiently large (equivalently  $t$  is sufficiently large),  $\hat{\sigma}_T$  will be *much* closer to one of  $\sigma_T$  or its reverse (nothing close to a tie can occur). To simplify notation, in what follows, we assume without loss of generality that  $\hat{\sigma}_T$  is closer to  $\sigma_T$ .

Fix  $(u, v) \in S^2$  such that  $0 \leq u < v \leq 1 - \mu$  and  $v > u + \epsilon \Delta_N$ . Choose any  $\kappa_1, \kappa_2$  such that  $\alpha < \kappa_1 < \kappa_2 < \mu$ , recall the definition of  $R_\sigma(\kappa_1)$  from equation (6.3),

$$\begin{aligned} \mathbb{E}[\hat{\Psi}_R(v) - \hat{\Psi}_R(u) | \mathcal{F}_T] &= \frac{1}{t} \sum_{r \in R} (w(v, r) - w(u, r)) \\ &= \frac{1}{t} \sum_{r \in R_{\sigma_T}(\kappa_1)} (w(v, r) - w(u, r)) + \frac{1}{t} \sum_{r \in R \setminus R_{\sigma_T}(\kappa_1)} (w(v, r) - w(u, r)). \end{aligned}$$

In what follows we will provide a lower bound for the the two summations appearing in  $\mathbb{E}[\hat{\Psi}_R(v) - \hat{\Psi}_R(u) | \mathcal{F}_T]$ . In fact there are two cases:

1.  $r \in R \setminus R_{\sigma_T}(\kappa_1)$  : Let  $K > 0$  be the constant from Assumption 2.1.

$$w(v, r) - w(u, r) \geq -K(v - u) \quad (6.21)$$

2.  $r \in R_{\sigma_T}(\kappa_1)$  : In this case,  $r = t_{(k)}$  for some  $k > (1 - \kappa_1)t$ , where  $t_{(k)}$ 's are the order statistics of the elements in  $T$ . By Lemma A.1,

$$r = t_{(k)} \geq \frac{k}{t} - \Delta_N > (1 - \kappa_1) - \Delta_N > (1 - \kappa_2) \quad \text{w.e.p.}$$

That is

$$r \in R_{\sigma_T}(\kappa_1) \Rightarrow 0 \leq u < v \leq 1 - \mu < 1 - \kappa_2 < r \leq 1 \quad \text{w.e.p.}$$

Therefore  $\min(|v - r|, |u - r|) \geq \mu - \kappa_2 > \frac{\beta - \alpha}{2}$  *w.e.p.* Applying Inequality (2.9) from Assumption 2.10, there exists some  $d_1 > 0$  such that

$$w(v, r) - w(u, r) \geq d_1(v - u) \quad \text{w.e.p.} \quad (6.22)$$

Combining Inequalities (6.21) and (6.22),

$$\mathbb{E}[\hat{\Psi}_R(v) - \hat{\Psi}_R(u) | \mathcal{F}_T] \geq \frac{d_1}{t} |R_{\sigma_T}(\kappa_1)| (v - u) - \frac{K}{t} |R \setminus R_{\sigma_T}(\kappa_1)| (v - u) \quad w.e.p.$$

Clearly  $|R_{\sigma_T}(\kappa_1)| \geq \kappa_1 t$ . Moreover, if  $r \in R \setminus R_{\sigma_T}(\kappa_1)$  then  $\hat{\sigma}_T(r) - \sigma_T(r) \geq (\kappa_1 - \alpha)t > 0$ , thus

$$\|\hat{\sigma}_T - \sigma_T\|_1 \geq (\kappa_1 - \alpha)t |R \setminus R_{\sigma_T}(\kappa_1)|. \quad (6.23)$$

Finally we have *w.e.p.*,

$$\begin{aligned} \mathbb{E}[\hat{\Psi}_R(v) - \hat{\Psi}_R(u) | \mathcal{F}_T] &\geq \frac{d_1}{t} |R_{\sigma_T}(\kappa_1)| (v - u) - \frac{K}{t} |R \setminus R_{\sigma_T}(\kappa_1)| (v - u) \\ &\stackrel{\text{Ineq (6.23)}}{\geq} \kappa_1 d_1 (v - u) - \frac{K}{(\kappa_1 - \alpha)t^2} \|\hat{\sigma}_T - \sigma_T\|_1 (v - u) \\ &\stackrel{\text{Ineq (6.20)}}{\geq} \left( \kappa_1 d_1 - \frac{CK}{\kappa_1 - \alpha} t^{-\eta} \right) (v - u) \\ &\geq \frac{\kappa_1 d_1}{2} (v - u) \\ &\geq \frac{\epsilon \kappa_1 d_1}{2} \Delta_N. \end{aligned}$$

Now, since  $\hat{\Psi}_R(v) - \hat{\Psi}_R(u)$  is a weighted sum of Bernoulli random variables sampled without uniformly without replacement. We can use Hoeffding's inequality for sampling without replacement [37, 38] to see that the conditional probability bound

$$\begin{aligned} \mathbb{P}[\hat{\Psi}_R(v) \geq \hat{\Psi}_R(u) | \mathcal{F}_T] &\geq 1 - \exp\left(-\frac{t^2}{2N} \left(\frac{\epsilon \kappa_1 d_1}{2}\right)^2 \Delta_\gamma^2(N)\right) \\ &\geq 1 - \exp(-\delta N \Delta_\gamma^2(N)) \\ &= 1 - \varepsilon_\gamma^\delta(N) \end{aligned}$$

holds *w.e.p.*<sup>5</sup> for some  $\delta > 0$ . The proof for  $\hat{\Psi}_L$  is identical to the proof of  $\hat{\Psi}_R$ . □

We now list some further facts about  $\hat{\Psi}_R, \hat{\Psi}_L$ :

**Proposition 6.5.** *Fix  $\epsilon > 0$ . The following events hold *w.e.p.*:*

1. For all  $v \in S$ ,

$$\begin{aligned} \{\hat{\Psi}_R(v) < c_R\} &\Rightarrow \{v \leq 1 - \mu\}, \\ \{\hat{\Psi}_L(v) < c_L\} &\Rightarrow \{v \geq \mu\}. \end{aligned}$$

2. For all pairs  $(u, v) \in S^2$ ,

$$\begin{aligned} \{\hat{\Psi}_R(u) < c_R, \hat{\Psi}_R(v) > c_R\} &\Rightarrow \{u < v \text{ or } |u - v| \leq \epsilon \Delta_N\}, \\ \{\hat{\Psi}_L(u) < c_L, \hat{\Psi}_L(v) > c_L\} &\Rightarrow \{u > v \text{ or } |u - v| \leq \epsilon \Delta_N\}. \end{aligned}$$

<sup>5</sup>When we say that an inequality holds *w.e.p.* we must specify what is being viewed as random. To readers unfamiliar with random measures, this might not be obvious in the preceding string of inequalities. We emphasize here that we are viewing the probability  $\mathbb{P}[\hat{\Psi}_R(j) < \hat{\Psi}_R(i) | \mathcal{F}_T]$  itself as an  $\mathcal{F}_T$ -measurable random variable (and we are viewing  $\mathbb{P}[\cdot | \mathcal{F}_T]$  as a random measure). Since  $\mathcal{F}_T$  is generated by the random choice of  $T$  and the many independent Bernoulli random variables contained in  $A|_T$ , it is  $T$  and  $A|_T$  that are viewed as random in this expression.

*Proof of Proposition 6.5.*

1. According to Lemmas 6.2 and 6.3, with extreme probability we have

$$\Psi_R(v) \leq \Psi_R(1 - \beta) + \Delta_N < \Psi_R(1 - \mu)$$

for all  $v \in S$  such that  $\hat{\Psi}_R(v) < c_R$ . By part 1 of Assumption 2.10, this implies that *w.e.p.* we have  $\max\{v \in S : \hat{\Psi}_R(v) < c_R\} < 1 - \mu$ . The proof for  $\hat{\Psi}_L$  is very similar.

2. Recall  $M$  from equation (6.17), define  $m = \inf_{x \in [0, 1 - \mu]} \Psi'_R(x)$ . Consider  $u, v \in S$  such that  $\hat{\Psi}_R(u) < c_R$  and  $\hat{\Psi}_R(v) > c_R$ . If  $u \geq v$  then according to the previous proposition  $0 \leq v \leq u \leq 1 - \mu$ . Moreover, applying Lemmas 6.2 and 6.3 with  $\epsilon' = \frac{\epsilon m}{2}$  we get

$$\Psi_R(1 - \beta) - \frac{\epsilon m}{2} \Delta_N \leq \Psi_R(v) \leq \Psi_R(u) \leq \Psi_R(1 - \beta) + \frac{\epsilon m}{2} \Delta_N \quad \textit{w.e.p.}$$

Therefore

$$|u - v| \leq \frac{1}{m} |\Psi_R(u) - \Psi_R(v)| \leq \Delta_N \quad \textit{w.e.p.}$$

The proof for  $\hat{\Psi}_L$  is very similar. □

Define the sets

$$\begin{aligned} \mathcal{E}_1 &= \{(u, v) \in S^2 : \hat{\Psi}_R(u), \hat{\Psi}_R(v) < c_R\}, \\ \mathcal{E}_2 &= \{(u, v) \in S^2 : \hat{\Psi}_L(u), \hat{\Psi}_L(v) < c_L\}, \\ \mathcal{E}_3 &= \{(u, v) \in S^2 : \hat{\Psi}_R(u) < c_R\}. \end{aligned}$$

Recall the antisymmetric function  $\hat{F} : S^2 \rightarrow \{-1, 1\}$  defined in equation (6.9) by

$$\hat{F}(u, v) = \begin{cases} 1 - 2 \cdot \mathbf{1}_{\hat{\Psi}_R(u) > \hat{\Psi}_R(v)} & \text{if } (u, v) \in \mathcal{E}_1, \\ 1 - 2 \cdot \mathbf{1}_{\hat{\Psi}_L(u) < \hat{\Psi}_L(v)} & \text{if } (u, v) \in \mathcal{E}_2, \cap \mathcal{E}_1^c, \\ 1 & \text{if } (u, v) \in \mathcal{E}_3 \cap \mathcal{E}_1^c \cap \mathcal{E}_2^c, \\ -1 & \text{if } (u, v) \in \mathcal{E}_1^c \cap \mathcal{E}_2^c \cap \mathcal{E}_3^c. \end{cases} \quad (6.24)$$

Define the antisymmetric function  $F : S^2 \rightarrow \{-1, 0, 1\}$  by  $F(u, v) = 1 - 2 \cdot \mathbf{1}_{u > v}$ . Note that the function  $\hat{F}$  appearing in equation (6.24) is intended to estimate the function  $F$ . Informally, Proposition 6.4 tells us that the first two branches of  $\hat{F}$  appearing in equation (6.24) would agree with  $F$  for all pairs of vertices  $u, v \in S \cap [0, 1 - \mu]$  if  $u$  and  $v$  are not *very close* and all pairs  $u, v \in S \cap [\mu, 1]$  if  $u$  and  $v$  are not *very close*. However, the function  $\hat{F}$  defined in equation (6.24) does not “know” if  $u, v$  are in  $[0, 1 - \mu]$  or  $[\mu, 1]$ ; it uses the four branches in its definition to guess. The following Lemma shows that the function  $\hat{F}(u, v)$  defined in equation (6.24) is equal to the function  $F(u, v)$  for all pairs of vertices  $u, v$  that are not *very close*.

**Lemma 6.6.** *Let  $\epsilon > 0$ . The event*

$$\{v > u + \epsilon \Delta_N\} \Rightarrow \{\hat{F}(u, v) = F(u, v)\} \quad (6.25)$$

*for all pairs  $(u, v) \in S^2$  holds w.e.p.*

*Proof of Lemma 6.6.* Let  $(u, v) \in S^2$  such that  $v > u + \epsilon\Delta_N$ . We prove equation (6.25) via a case-by-case analysis of the performance of  $\hat{F}$  in its four branches. In fact we have four cases:

1.  $\mathcal{E}_1$  **holds**: By the first part of Proposition 6.5, we have  $u, v < 1 - \mu$  *w.e.p.* Applying Proposition 6.4 completes the proof of equation (6.25) in this case.
2.  $\mathcal{E}_2 \cap \mathcal{E}_1^c$  **holds**: By the first part of Proposition 6.5, we have  $u, v > \mu$  *w.e.p.* Applying Proposition 6.4 completes the proof in this case.
3.  $\mathcal{E}_3 \cap \mathcal{E}_1^c \cap \mathcal{E}_2^c$  **holds**: By the second part of Proposition 6.5, there are two possibilities. Either  $\hat{F}$  makes the correct decision, or  $|u - v| \leq \epsilon\Delta_N$  *w.e.p.*
4.  $\mathcal{E}_1^c \cap \mathcal{E}_2^c \cap \mathcal{E}_3^c$  **holds**: Again by the second part of Proposition 6.5, there are two possibilities. Either  $\hat{F}$  makes the correct decision, or  $|u - v| \leq \epsilon\Delta_N$  *w.e.p.* □

Similarly, recall the function  $\hat{f}$  defined in equation (6.10). The following Lemma shows that the function  $\hat{f}$  makes the correct ordering “guess” almost all the time; that is for  $(u, v) \in S^2$  such that  $u$  and  $v$  are not *very close* we have  $(\hat{f}(v) - \hat{f}(u))(v - u) > 0$ .

**Lemma 6.7.** *Fix  $\epsilon > 0$ . The event that*

$$\{v > u + \epsilon\Delta_N\} \Rightarrow \{\hat{f}(v) > \hat{f}(u)\}$$

*holds for all pairs  $(u, v) \in S^2$  occurs w.e.p.*

*Proof Lemma 6.7.* Fix  $(u, v) \in S^2$  with  $v > u + \epsilon\Delta_N$ . Define the set

$$\mathcal{A} = \left\{ r \in V : \min(|u - r|, |v - r|) > \frac{\epsilon}{8}\Delta_N \right\}.$$

By Lemma 6.6, for all  $r \in \mathcal{A}$

$$\hat{F}(r, v) = F(r, v) \text{ and } \hat{F}(r, u) = F(r, u) \quad \textit{w.e.p.}$$

Therefore we have *w.e.p.*

$$\begin{aligned} \hat{f}(v) - \hat{f}(u) &= \sum_{r \in V} \mathbf{1}_{r \in S} \left( \hat{F}(r, v) - \hat{F}(r, u) \right) \\ &= \sum_{r \in \mathcal{A}} \mathbf{1}_{r \in S} (F(r, v) - F(r, u)) + \sum_{r \in \mathcal{A}^c} \mathbf{1}_{r \in S} \left( \hat{F}(r, v) - \hat{F}(r, u) \right) \\ &\geq 2|\{r \in S : u < r < v\}| - 2|S \cap \mathcal{A}^c| \\ &= 2(A - B), \end{aligned}$$

where

$$A = \sum_{k \in E_1} \mathbf{1}_{v_k \in S}, \quad B = \sum_{k \in E_2} \mathbf{1}_{v_k \in S},$$

with  $E_1 = \{k \in [N] : Nu < k < Nv\}$  and  $E_2 = \{k \in [N] : \min(|k - Nu|, |k - Nv|) \leq \lfloor \frac{\epsilon}{8}N\Delta_N \rfloor\}$ . Note that,  $A$  and  $B$  appearing in this bound are sums of Bernoulli random variables sampled without replacement. That is  $A, B$  are hypergeometric random variables, with  $A \sim \mathcal{HG}(N; Nv - Nu - 1; s)$  and  $B \sim \mathcal{HG}(N; 4 \lfloor \frac{\epsilon}{8}N\Delta_N \rfloor + 2; s)$ .

We have

$$\mathbb{E}[A - B] = \frac{s}{N} \left( N(v - u) - 4 \left\lfloor \frac{\epsilon}{8} N \Delta_N \right\rfloor - 3 \right) > \frac{\epsilon}{10} N \Delta_N \quad w.e.p.$$

Thus, by the usual analogue to Hoeffding's inequality for hypergeometric random variables [37, 38], we have

$$\hat{f}(v) - \hat{f}(u) > 0 \quad w.e.p. \quad (6.26)$$

□

We are now ready to prove Theorem 6.1.

*Proof of Theorem 6.1.* Let  $s_{(k)}$ 's be the order statistics of the set  $S$ . Fix  $1 \leq k \leq s$ , recall that

$$\hat{\sigma}_S(s_{(k)}) = \left| \left\{ p \in [s] : \hat{f}(s_{(p)}) \leq \hat{f}(s_{(k)}) \right\} \right| \quad \text{and} \quad \sigma_S(s_{(k)}) = k.$$

For any integer  $p \leq k - \frac{\epsilon}{2} N \Delta_N$  we have  $s_{(p)} \leq s_{(k)} - \frac{\epsilon}{2} \Delta_N$ , thus by Lemma 6.7

$$\forall p \leq \left\lfloor k - \frac{\epsilon}{2} N \Delta_N \right\rfloor : \hat{f}(s_{(p)}) < \hat{f}(s_{(k)}) \quad w.e.p.$$

Therefore

$$\hat{\sigma}_S(s_{(k)}) \geq \left\lfloor k - \frac{\epsilon}{2} N \Delta_N \right\rfloor \geq k - \epsilon N \Delta_N \quad w.e.p.$$

Similarly for any  $p \geq k + \frac{\epsilon}{2} N \Delta_N$  we have  $s_{(p)} \geq s_{(k)} + \frac{\epsilon}{2} \Delta_N$ , thus by Lemma 6.7

$$\forall p \geq \left\lceil k + \frac{\epsilon}{2} N \Delta_N \right\rceil : \hat{f}(s_{(p)}) > \hat{f}(s_{(k)}) \quad w.e.p.$$

Therefore

$$\hat{\sigma}_S(s_{(k)}) \leq \left\lceil k + \frac{\epsilon}{2} N \Delta_N \right\rceil \leq k + \epsilon N \Delta_N \quad w.e.p.$$

Finally,

$$(\forall v \in S) : |\hat{\sigma}_S(v) - \sigma_S(v)| \leq \epsilon N \Delta_N = \epsilon \sqrt{N \log^\gamma(N)} \quad w.e.p.,$$

as required. □

#### 6.4. Proof of Theorem 2.13 and analysis of Algorithm 3

We now prove the main bounds on the objects appearing in Algorithm 3, completing the proof of Theorem 2.13. Throughout this section, we use notation from Algorithm 3 and make the same assumptions as in the statement of Theorem 2.13. By Theorem 6.1,

$$\|\hat{\sigma}_j - \sigma_{S_j}\|_\infty \leq \frac{\epsilon}{12} \sqrt{N \log^\gamma(N)} \quad (6.27)$$

holds for all  $j \in \{1, 2, 3\}$  *w.e.p.*

We next discuss Step 3 of the algorithm. Recall the definition of *aligned* and *closely aligned* from Inequality (6.2). Note that  $\{\hat{\sigma}_j\}_{j \in \{1,2,3\}}$  were computed in separate calls to Algorithm 2 and so they may not be aligned - that is, some  $\hat{\sigma}_j$  may be aligned with  $\sigma_{S_j}$  while others may be aligned more with its reverse  $\text{rev}(\sigma_{S_j})$ . When  $N$  is large enough all the  $\sigma_{S_j}$ 's are closely aligned with  $id_N$ . Moreover, when Inequality (6.27) holds and  $N$  sufficiently large,  $\hat{\sigma}_j$  will either be closely aligned with  $\sigma_{S_j}$  or with  $\text{rev}(\sigma_{S_j})$ . That is,  $\hat{\sigma}_j$  will be closely aligned with  $id_N$  or with its reverse for all  $j \in \{1, 2, 3\}$ . For the same reason, for  $N$  sufficiently large  $\hat{\sigma}_{\text{ref}}$  will either be closely aligned with  $id_N$  or with  $\text{rev}(id_N)$ . We assume without loss of generality that  $\hat{\sigma}_{\text{ref}}$  is closely aligned with  $id_N$ . Finally, note that *w.e.p.*,  $|S_j \cap S'_2| \geq 2$  for all  $j \in \{1, 2, 3\}$ . When all of these events occur, for each  $j \in \{1, 2, 3\}$ , one of the following two cases must occur:

1.  $\hat{\sigma}_j$  is closely aligned with  $id_N$ . Since  $|S_j \cap S'_2| \geq 2$  and  $\hat{\sigma}_j, \hat{\sigma}_{\text{ref}}$  are both closely aligned to  $id_N$ , they must also be closely aligned to each other. Thus, in this case,  $\hat{\sigma}_j$  is *not* reversed in step 3 of the algorithm and *remains* closely aligned to  $id_N$  in step 4.
2.  $\hat{\sigma}_j$  is closely aligned with  $\text{rev}(id_N)$ . Since  $|S_j \cap S'_2| \geq 2$  and  $\text{rev}(\hat{\sigma}_j), \hat{\sigma}_{\text{ref}}$  are both closely aligned to  $id_N$ , they must also be aligned to each other. Thus, in this case,  $\hat{\sigma}_j$  is reversed in step 3 of the algorithm and *becomes* closely aligned to  $id_N$  in step 4.

Thus, in both cases, by step 4 the ordering  $\hat{\sigma}_j$  is closely aligned to  $id_N$  for all  $j \in \{1, 2, 3\}$  *w.e.p.*

Denote by  $s_1^{(k)}, s_2^{(k)}$  and  $s_3^{(k)}$  the order statistics of the sets  $S_1, S_2$  and  $S_3$  respectively. Then by equation (6.1) we have

$$\begin{aligned}
 \|\hat{\sigma} - id_N\|_\infty &= \|\hat{\sigma}^{-1} - id_N\|_\infty \\
 &= N \max_{j \in \{1,2,3\}} \max_{k \in \{1, \dots, n+1\}} |\hat{\sigma}_j^{-1}(k) - v_{3k+j-3}| \\
 &\leq N \max_{j \in \{1,2,3\}} \max_{k \in \{1, \dots, n+1\}} |\hat{\sigma}_j^{-1}(k) - v_{3k}| + 2 \\
 &= \max_{j \in \{1,2,3\}} \max_{k \in \{1, \dots, n+1\}} |3\hat{\sigma}_j(s_j^{(k)}) - Ns_j^{(k)}| + 2 \\
 &\leq 3 \max_{j \in \{1,2,3\}} \|\hat{\sigma}_j - \sigma_{S_j}\|_\infty + N \max_{j \in \{1,2,3\}} \max_{k \in \{1, \dots, n+1\}} |s_j^{(k)} - v_{3k}| + 2.
 \end{aligned}$$

By Lemma A.1,

$$|s_j^{(k)} - v_{3k}| \leq \frac{\epsilon}{4} \Delta_N \quad \textit{w.e.p.}, \tag{6.28}$$

for all  $j \in \{1, 2, 3\}$  and  $k \in \{1, \dots, n+1\}$  such that  $3k+j \leq N$ . When  $N$  is large enough, and when inequalities (6.27) and (6.28) hold, we get

$$\|\hat{\sigma} - id_N\|_\infty \leq \epsilon \sqrt{N \log^7(N)} \quad \textit{w.e.p.}$$

which completes the proof.

**Remark 6.8** (Learning  $\alpha$  and  $\beta$ ). Algorithm 2 takes parameters  $\alpha, \beta$  as given. However, our results only apply if these parameters satisfy Assumption 2.10. In Theorem 1.3, we state values of  $(\alpha, \beta)$  for which the graphons satisfying Assumption 1.2 will also satisfy Assumption 2.10, and so there are no difficulties in that setting.

However, for other graphons, it may not be obvious how to choose  $\alpha, \beta$ . We note here that, when graphons satisfy a slightly strengthened version of the assumption for some pair  $\alpha, \beta$ , it is not difficult to learn this pair quickly. We now give details.

Note that the functions  $\Psi_R = \Psi_R^{(\alpha)}, \Psi_L = \Psi_L^{(\alpha)}$  depend on the parameter  $\alpha$  (though this dependence is suppressed in our notation). Recall that  $\Psi_R, \Psi_L$  are continuous functions. If  $w$  satisfies Inequality (2.8) for some

$\alpha < \beta$ , then there exists some  $\delta > 0$  and some other pair  $\alpha', \beta'$  so that it satisfies the slightly strengthened bound

$$\inf_{x \in [1-\alpha'-2\delta, 1]} \Psi_R^{(\alpha')}(x) > \Psi_R^{(\alpha')}(1 - \beta') + 2\delta, \quad \inf_{x \in [0, \alpha'+2\delta]} \Psi_L^{(\alpha')}(x) > \Psi_L^{(\alpha')}(\beta') + 2\delta, \quad (6.29)$$

By Lemma 6.2 the estimates  $\hat{\Psi}_R, \hat{\Psi}_L$  calculated in equations (6.5), (6.6) have error  $\Delta_N$  going to 0 with  $N$ . Thus, for all  $N$  sufficiently large, with extreme probability any pair  $(\alpha', \beta')$  satisfying (6.29) will also satisfy

$$\inf_{x \in [1-\alpha'-\delta, 1]} \hat{\Psi}_R^{(\alpha')}(x) > \hat{\Psi}_R^{(\alpha')}(1 - \beta') + \delta, \quad \inf_{x \in [0, \alpha'+\delta]} \hat{\Psi}_L^{(\alpha')}(x) > \hat{\Psi}_L^{(\alpha')}(\beta') + \delta. \quad (6.30)$$

Applying Lemma 6.2 again, any pair satisfying the Inequality (6.30) will also satisfy Inequality (2.8).

Thus, if a graphon satisfies (6.29) for some  $\delta > 0$  and some pair  $(\alpha, \beta)$ , we can find the pair  $(\alpha', \beta')$  by using Inequality (6.30) as a test. Since the bound in Lemma 6.2 is a bound on the *maximum* error, checking *many* pairs  $(\alpha', \beta')$  from a single observed graph does not present any obstacles.

## 7. APPLICATION

We show that Theorem 1.3 follows from Theorems 2.8 and 2.13. To do so, it is enough to check that any graphon that is “nice” in the sense of Definition 1.2 will also satisfy Assumptions 2.1 through 2.10.

We now check Assumptions 2.1 through 2.10 in order.

**Assumption 2.1:** Define  $R_{\max} = \sup_{r \in (0, 1]} R'(r) < \infty$ . We calculate

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

**Assumption 2.2:** We have  $\frac{\partial w(x, y)}{\partial x} = R'(y)$  for  $y \neq x$ , so this follows immediately from Definition 1.2.

**Assumption 2.4:** We calculate

$$\begin{aligned} d'(x) &= \frac{d}{dx} \int_0^1 R(|x - y|) dy \\ &= \frac{d}{dx} \left( \int_0^x R(y) dy + \int_0^{1-x} R(y) dy \right) \\ &= R(x) - R(1 - x). \end{aligned}$$

Thus  $d'(x)$  exists for all  $x \in [0, 1]$ . Since  $R$  is monotone,  $d'(x) > 0$  for  $x \in [0, 0.5)$  and  $d'(x) < 0$  for  $x \in (0.5, 1]$ , with  $d'(x) = 0$  only for  $x = 0.5$ .

**Assumption 2.7:** This is given in [33], with the function  $h$  in that reference replacing the function  $R$  in this paper.

**Assumption 2.10:** To prove Inequality (2.8), we have

$$\begin{aligned} \inf_{x \in [0, 0.05]} \Psi_L(x) &= \inf_{x \in [0, 0.05]} \int_0^{0.05} w(x, y) dy \\ &= \inf_{x \in [0, 0.05]} \int_0^{0.05} R(|x - y|) dy. \end{aligned}$$

Since  $R$  is monotone, this infimum is achieved at  $x = 0$  and  $x = 0.05$ . Continuing the calculation,

$$\begin{aligned} \inf_{x \in [0, 0.05]} \Psi_L(x) &= \int_0^{0.05} R(y) dy \\ &> \int_{0.26}^{0.31} R(y) dy \\ &= \Psi_L(0.31), \end{aligned}$$

where the strict inequality in the second line comes from the assumption that  $R'$  is continuous and strictly positive. This proves one of the four inequalities in the first part Assumption 2.10; the remainder are essentially the same.

Finally, we check Inequality (2.9). Consider a triple  $x, y, z$ . Assume WLOG that  $d_x \equiv |x - z| > d_y \equiv |y - z|$ . Define  $D_{\min} = \inf_{a \in (0, 1)} -R'(a) > 0$ . We have

$$\begin{aligned} |w(x, z) - w(y, z)| &= |R(|x - z|) - R(|y - z|)| \\ &= |R(d_x) - R(d_y)| \\ &\geq |d_x - d_y| \inf_{a \in (d_x, d_y)} |R'(a)| \\ &\geq D_{\min} |d_x - d_y|, \end{aligned}$$

where the first inequality follows immediately from applying Gronwall's inequality. This proves the desired bound.

## 8. DISCUSSION

In this paper, we have written down a simple seriation algorithm and shown that, for some fairly large class of model graphons, the result of the algorithm is an estimated permutation  $\hat{\sigma}$  with sup-norm error of roughly  $O(\sqrt{N})$ . It is natural to ask whether either the rates of convergence or our assumptions about the generating graphon are optimal.

The second question seems to have an easy answer: we are quite sure that our conditions are not optimal, even for the type of analysis we use. In particular, our main assumptions on  $w$  imply that all vertices have a fairly non-negligible chance of being connected (more precisely, our main assumptions imply that  $w(x, y) > 0$  for all  $x, y \in [0, 1]$ ). We have already seen in Remark 2.3 that this property is not satisfied for many natural graphons, and also that it can be substantially weakened.

We hypothesize that this global connection property can be replaced by a much weaker assumption, saying roughly that some *power* of the adjacency matrix should be globally connected. The main difficulties in carrying this out are different for our two main results:

- The main difficulty in extending Theorem 2.8 is carrying through the arguments in Section 3, which show that the Fiedler eigenspace has dimension 1 and that the Fiedler function is monotone. As discussed in Remark 2.3, this can sometimes be checked directly for particularly simple graphons. We also expect that certain generic techniques, such as perturbation analysis, can be used to extend this to small nonparametric families of graphons. However, we do not know how to get rid of this assumption in any great generality. While we believe that our assumptions could be weakened, we don't believe that spectral seriation will give good results for all statistically-tractable seriation problems – see *e.g.* the discussion in [26].
- We believe that improving Theorem 2.8 relies mostly on improving the analysis for a fixed algorithm. By contrast, it is clear that the strongest versions of Theorem 2.13 require some changes to Algorithm 3. In particular, Algorithm 3 will not generally give a sensible answer without the global connectedness property.

This occurs because the functions  $\psi_R(i), \psi_L(i)$  may both be 0 for many  $i \in [N]$ . In [19], we introduce a “local” quantity  $R(i, j)$  that is analogous to the “global” quantity  $R$  used in the present paper. We suspect that a similar approach would work in the present context.

We don’t know the optimal rate of convergence for this problem in very many metrics, but we give a few related results largely drawn from our companion paper [19]. Note that this paper studies a slightly different sampling scheme, but very similar arguments apply:

1. Under slightly different model conditions, the iterated post-processing algorithm of [19] can obtain error of  $O(N^\epsilon)$  for any fixed  $\epsilon > 0$ , so the rate  $N^{0.5}$  is certainly not optimal for all classes of graphon models.
2. On the other hand, we might ask to estimate the latent positions  $v_1, \dots, v_N$  of vertices in addition to the latent permutation  $\sigma$ . The argument in Section 5 of [19] says that any such estimate must suffer an error of at least  $\Omega(N^{-0.5})$  for any sufficiently large class of graphon models. Since the natural estimate

$$\hat{v}_i = \frac{\hat{\sigma}(i)}{N}$$

is so similar to  $\hat{\sigma}$  itself, this suggests that one can’t do much better in general. This rough conclusion is also supported by similar results in [23, 26].

Several other papers have recently studied the optimal rate of convergence for “the” noisy seriation problem, including [39] and [24]. Their models and versions of the noisy seriation problem are different enough from ours that it is not possible to directly compare the results. We leave the question of comparing these notions of noisy seriation to future work. Finally, we note that we obtain a bound in the strong sup-norm only for the post-processed results, not for the spectral initialization. This do not appear to be fundamental obstacles to getting sup-norm bounds for spectral seriation itself under strong assumptions; see *e.g.* [40].

#### DATA AVAILABILITY STATEMENT

No new data or code were created or analyzed in this study.

#### REFERENCES

- [1] W.S. Robinson, A method for chronologically ordering archaeological deposits. *Am. Antiq.* **16** (1951) 293–301.
- [2] W.F. Petrie, Sequences in prehistoric remains. *J. Anthropol. Inst. Great Britain Ireland* **29** (1899) 295–301.
- [3] I. Liiv, Seriation and matrix reordering methods: an historical overview. *Statist. Anal. Data Mining* **3** (2010) 70–91.
- [4] D.G. Corneil, A simple 3-sweep LBFS algorithm for the recognition of unit interval graphs. *Discrete Appl. Math.* **138** (2004) 371–379.
- [5] J.E. Atkins, E.G. Boman and B. Hendrickson, A spectral algorithm for seriation and the consecutive ones problem. *SIAM J. Comput.* **28** (1998) 297–310.
- [6] H. Chuangpishit, M. Ghandehari, M. Hurshman, J. Janssen and N. Kalyaniwalla, Linear embeddings of graphs and graph limits. *J. Combinat. Theory Ser. B* **113** (2015) 162–184.
- [7] L. Lovász, Large Networks and Graph Limits, Vol. 60. American Mathematical Society (2012).
- [8] V. Bagaria, J. Ding, D. Tse, Y. Wu and J. Xu, Hidden Hamiltonian cycle recovery via linear programming. *Oper. Res.* **68** (2020) 53–70.
- [9] T. Cai, T. Liang and A. Rakhlin, On detection and structural reconstruction of small-world random networks. *IEEE Trans. Netw. Sci. Eng.* **4** (2017) 165–176.
- [10] J. Ding, Y. Wu, J. Xu and D. Yang, Consistent recovery threshold of hidden nearest neighbor graphs. *Proceedings of Thirty Third Conference on Learning Theory*, Vol. 125 (2020) 1540–1553.
- [11] Y. Issartel, C. Giraud and N. Verzelen, Minimax optimal seriation in polynomial time (2024) <https://www.semanticscholar.org/paper/Minimax-optimal-seriation-in-polynomial-time-Issartel-Giraud/ba34785e6579da054149358453b7f368b61d929d>.

- [12] W.E. Donath and A.J. Hoffman, Lower bounds for the partitioning of graphs, in Selected Papers Of Alan J. Hoffman: With Commentary. World Scientific (2003) 437–442.
- [13] U. Von Luxburg, A tutorial on spectral clustering. *Statist. Comput.* **17** (2007) 395–416.
- [14] M. Avella-Medina, F. Parise, M. Schaub and S. Segarra, Centrality measures for graphons: accounting for uncertainty in networks. *IEEE Trans. Netw. Sci. Eng.* (2018).
- [15] S. Chatterjee, Matrix estimation by universal singular value thresholding. *Ann. Statist.* **43** (2015) 177–214.
- [16] F. Krzakala, C. Moore, E. Mossel, J. Neeman, A. Sly, L. Zdeborová and P. Zhang, Spectral redemption in clustering sparse networks. *Proc. Natl. Acad. Sci. U.S.A.* **110** (2013) 20935–20940.
- [17] U. Von Luxburg, M. Belkin and O. Bousquet, Consistency of spectral clustering. *Ann. Statist.* **36** (2008) 555–586.
- [18] A. Daniely, N. Linial and M. Saks, Clustering is difficult only when it does not matter. preprint arXiv:1205.4891 (2012).
- [19] J. Janssen and A. Smith, Reconstruction of line-embeddings of graphons. *Electron. J. Statist.* **16** (2022) 331–407.
- [20] E. Abbe, Community detection and stochastic block models: recent developments. *J. Mach. Learn. Res.* **18** (2018) 1–86.
- [21] F. Fogel, A. d’Aspremont and M. Vojnovic, Serialrank: spectral ranking using seriation, in Advances in Neural Information Processing Systems (2014) 900–908.
- [22] I. Rocha, J. Janssen and N. Kalyaniwalla, Recovering the structure of random linear graphs. *Linear Algebra Appl.* **557** (2018) 234–264.
- [23] C. Giraud, Y. Issartel and N. Verzelen, Localization in 1D non-parametric latent space models from pairwise affinities. *Electron. J. Statist.* **17** (2023) 1587–1662.
- [24] N. Flammarion, C. Mao and P. Rigollet, Optimal rates of statistical seriation. *Bernoulli* **25** (2019) 623–653.
- [25] A. Recanatì, N. Servant, J.-P. Vert and A. d’Aspremont, Robust Seriation and Applications to Cancer Genomics. arXiv preprint arXiv:1806.00664 (2018).
- [26] T.T. Cai and R. Ma, Matrix reordering for noisy disordered matrices: optimality and computationally efficient algorithms. *IEEE Trans. Inform. Theory* **70** (2024) 509–531.
- [27] C. Mao, J. Weed and P. Rigollet, Minimax rates and efficient algorithms for noisy sorting. *ALT* (2019) <https://proceedings.mlr.press/v83/mao18a.html>.
- [28] D. Glasscock, What is... a Graphon? *Notices AMS* **62** (2015).
- [29] M. Reed, Methods of Modern Mathematical Physics: Functional Analysis. Elsevier (2012).
- [30] D.E. Edmunds, A. Potter and C.A. Stuart, Non-compact positive operators. *Proc. R. Soc. Lond. A* **328** (1972) 67–81.
- [31] K. Gustafson and J. Weidmann, On the essential spectrum. *J. Math. Anal. Appl.* **25** (1969) 121–127.
- [32] G. Teschl, Mathematical methods in quantum mechanics. *Grad. Stud. Math.* **99** (2009) 106.
- [33] I. Pinelis, Infimum upper bound (answer). MathOverflow. URL:<https://mathoverflow.net/questions/324561/infimum-upper-bound> (visited on 2021-09-10).
- [34] F. Chung and M. Radcliffe, On the spectra of general random graphs. *Electron. J. Combinatorics* **18** (2011) 215.
- [35] C. Davis and W.M. Kahan, The rotation of eigenvectors by a perturbation. *SIAM J. Numer. Anal.* **7** (1970) 1–46.
- [36] P. Diaconis and R.L. Graham, Spearman’s footrule as a measure of disarray. *J. Roy. Statist. Soc. Ser. B (Methodological)* **39** (1977) 262–268.
- [37] E. Greene and J.A. Wellner, Exponential bounds for the hypergeometric distribution. *Bernoulli* **23** (2017) 1911.
- [38] R.J. Serfling, Probability inequalities for the sum in sampling without replacement. *Ann. Statist.* **2018** (1974) 39–48.
- [39] C. Mao, A. Pananjady and M.J. Wainwright, Breaking the  $1/\sqrt{n}$  barrier: faster rates for permutation-based models in polynomial time, in *Proceedings of the 31st Conference on Learning Theory*, edited by S. Bubeck, V. Perchet and P. Rigollet. Proceedings of Machine Learning Research Vol. 75 (2018) 2037–2042.
- [40] E. Abbe, J. Fan, K. Wang and Y. Zhong, Entrywise eigenvector analysis of random matrices with low expected rank. *Ann. Statist.* **48** (2020) 1452–1474.



**Please help to maintain this journal in open access!**

This journal is currently published in open access under the Subscribe to Open model (S2O). We are thankful to our subscribers and supporters for making it possible to publish this journal in open access in the current year, free of charge for authors and readers.

Check with your library that it subscribes to the journal, or consider making a personal donation to the S2O programme by contacting [subscribers@edpsciences.org](mailto:subscribers@edpsciences.org).

More information, including a list of supporters and financial transparency reports, is available at <https://edpsciences.org/en/subscribe-to-open-s2o>.

APPENDIX A

**Lemma A.1.** *Let  $V = \{v_1, v_2, \dots, v_N\}$  with  $v_i = i/N$  for all  $i$ . Let  $S$  be a subset of  $V$  chosen uniformly at random and with deterministic size  $m = |S|$  such that  $\lfloor \frac{N}{3} \rfloor \leq m \leq \lfloor \frac{N}{3} \rfloor + 1$ . Define  $s_{(1)} < s_{(2)} < \dots < s_{(m)}$  to be the order statistics of the elements of  $S$ . Then for each  $1 \leq k \leq m$  and for any  $c > 0$ , there exists a constant  $\delta > 0$  and an integer  $N_0$  such that*

$$\mathbb{P} \left( \left| s_{(k)} - \frac{k}{m} \right| > c\Delta_N \right) \leq \varepsilon_\gamma^\delta(N)$$

for all  $N \geq N_0$ . In other words for each  $1 \leq k \leq m$  and  $c > 0$  the event  $\left| s_{(k)} - \frac{k}{m} \right| \leq c\Delta_N$  holds w.e.p.

*Proof of Lemma A.1.* Let  $1 \leq k \leq N$  and  $k \leq j \leq N - m + k$ . Then  $s_{(k)} = v_j$  if and only if there is exactly  $k - 1$  elements from  $S$  less than  $v_j$  and exactly  $m - k$  elements from  $S$  bigger than  $v_j$ . Therefore

$$\mathbb{P} (s_{(k)} = v_j) = \frac{\binom{j-1}{k-1} \binom{N-j}{m-k}}{\binom{N}{m}}.$$

Note that the distribution of  $s_{(k)}$  looks very similar to a negative hypergeometric distribution. In fact one can easily check that  $Ns_{(k)} - k \sim \mathcal{NHG}(N; N - m; k)$ .

The paper [38] provides concentration inequalities for sampling without replacement. According to [37, 38], for any  $\lambda > 0$  we have

$$\mathbb{P} (|s_{(k)} - \mathbb{E}[s_{(k)}]| > \lambda) \leq 2 \exp \left( \frac{-2N^2\lambda^2}{m(1 - f_m^*)} \right)$$

with  $f_m^* = (m - 1)/N$ . We can check that  $1/4 < f_m^* \leq 1/3$  for all  $N$  sufficiently large. Thus,

$$\mathbb{P} (|s_{(k)} - \mathbb{E}[s_{(k)}]| > \lambda) \leq 2 \exp \left( \frac{-8N^2\lambda^2}{3m} \right).$$

Let  $c > 0$ , we have

$$\begin{aligned} \left| s_{(k)} - \frac{k}{m} \right| > c\Delta_N &\Rightarrow |s_{(k)} - \mathbb{E}[s_{(k)}]| > c\Delta_N - \left| \frac{k}{m} - \mathbb{E}[s_{(k)}] \right| \\ &\Rightarrow |s_{(k)} - \mathbb{E}[s_{(k)}]| > c\Delta_N - \frac{(N - m)k}{Nm(m + 1)} \\ &\Rightarrow |s_{(k)} - \mathbb{E}[s_{(k)}]| > c\Delta_N - \frac{2}{N}. \end{aligned}$$

For  $N$  large enough so that  $4 < cN\Delta_N$ ,

$$|s_{(k)} - \frac{k}{m}| > c\Delta_N \Rightarrow |s_{(k)} - \mathbb{E}[s_{(k)}]| > \frac{c}{2}\Delta_N.$$

Therefore

$$\begin{aligned} \mathbb{P}\left(|s_{(k)} - \frac{k}{m}| > c\Delta_N\right) &\leq \mathbb{P}\left(|s_{(k)} - \mathbb{E}[s_{(k)}]| > \frac{c}{2}\Delta_N\right) \\ &\leq 2 \exp\left(\frac{-2c^2 N^2 \Delta_N^2}{3m}\right) \\ &\leq 2 \exp(-c^2 N \Delta_N^2) \\ &\leq \varepsilon_\gamma^\delta(N) \end{aligned}$$

for some  $0 < \delta < c^2$ .

□