Random Geometric Graphs on Euclidean Balls

Ernesto Araya Valdivia
Laboratoire de Mathématiques d’Orsay (LMO)
Université Paris-Saclay, 91405 Orsay Cedex, France

Abstract: We consider a latent space model for random graphs where a node $i$ is associated to a random latent point $X_i$ on the Euclidean unit ball. The probability that an edge exists between two nodes is determined by a “link” function, which corresponds to a dot product kernel. For a given class $\mathcal{F}$ of spherically symmetric distributions for $X_i$, we consider two estimation problems: latent norm recovery and latent Gram matrix estimation. We construct an estimator for the latent norms based on the degree of the nodes of an observed graph in the case of the model where the edge probability is given by $f((X_i, X_j)) = \mathbb{1}_{\langle X_i, X_j \rangle \geq \tau}$, where $0 < \tau < 1$. We introduce an estimator for the Gram matrix based on the eigenvectors of observed graph and we establish Frobenius type guarantee for the error, provided that the link function is sufficiently regular in the Sobolev sense and that a spectral-gap-type condition holds. We prove that for certain link functions, the model considered here generates graphs with degree distribution that have tails with a power-law-type distribution, which can be seen as an advantage of the model presented here with respect to the classic Random Geometric Graph model on the Euclidean sphere. We illustrate our results with numerical experiments.

MSC 2010 subject classifications: Primary 68Q32; secondary 60F99, 68T01.
Keywords and phrases: Random geometric graph, Graphon spectrum, Gram matrix estimation, Latent space model.

1. Introduction

Given the ubiquity of network structured databases, the task of extracting information from them has become an important topic within many scientific communities, including statistics and machine learning. This has gone in hand with the development, mainly in the last decade, of powerful tools of graph theory, such as the graphon theory [Lovász and Szegedy, 2006, Borgs et al., 2008b, Borgs et al., 2008a], which describes the asymptotic behavior of large dense graphs.

In this paper we will focus on extracting information from a single observation of a graph, which we assume generated from a parametric family of models, with latent space structure, which we will call random geometric graphs (RGG) on the Euclidean ball $B^d = \{x \in \mathbb{R}^d : \|x\| \leq 1\}$. The model we will consider here has similarities not only with the random geometric graph model on the sphere and its generalizations, considered for example in [Bubeck et al., 2016, De Castro et al., 2020], but also with the random dot product graph model (RDPG) [Athreya et al., 2018, Sussman et al., 2014]. Indeed, one of our goals is to show that the random graph model presented here is flexible enough to generate graphs that have a degree profile distributed according to a power-law type distribution, while maintain some of the structural qualities that make it well-suited for statistical inference.

We will consider a particular instance of the $W$-random graph model for dense graphs [Lovasz, 2012, Ch.10], where a kernel function $W$ defines the probability of connection between two latent points. Similar to the context treated in [Araya and De Castro, 2019], we will consider $W$ to be a dot product kernel, but here the ambient space will be the Euclidean ball, instead of the unit sphere. More specifically, we will consider that each node of a
graph is associated with a randomly placed latent point in $B^d$ (in an i.i.d manner) according to a probability distribution that belongs to a parametric family of spherically symmetric distribution, that we will call $\mathcal{F}$. The main difference with the spherical case is that when considering $B^d$ as the ambient space it is not only the angle between the latent points which determines the probability of connection between two nodes, but also their norm. This offer more flexibility in the degree distribution of the generated graphs, which in the spherical case is concentrated around a single value, at least when the latent points are distribution according to the uniform distribution (which is the only spherically symmetric probability measure). In particular, for certain link functions, we will show that the degree sequence exhibits a power-law type distribution. To best of our knowledge, there is no standard definition of power law type distribution in the graphon literature. We will introduce a notion of power-law distributions in this context based on the (normalized) degree function $d_W(\cdot)$ defined on graphon [Lovasz, 2012, Sec.7.1].

We discuss two problems of estimation of latent information on this model. We first study possibility of estimation of the latent norm from the observed adjacency matrix, in the threshold graphon model, that is when the link function (or graphon) is of the form $f((x, y)) = 1_{\langle x, y \rangle \geq \tau}$, for a $\tau > 0$ and $x, y \in B^d$. In this model, two nodes will be connected if their latent points have inner product larger than $\tau$. We propose an estimator for the norm of the latent points based on the degree of the correspondent latent point. We prove the consistency of the estimator and illustrate its performance by simulations.

We next study the problem of estimating the Gram matrix of the latent points for the RGG model on $B^d$, proposing an estimator which is based on a set of eigenvectors of the observed adjacency matrix, which extends the spectral approach developed in the spherical case [Araya and De Castro, 2019]. Our main assumption is related to the spectral gap between certain eigenvalues of the integral operator associated to the link function. This type of assumptions have been used before in the literature, mainly in the context of matrix estimation and manifold learning, often because some version of the Davis-Kahan $\sin \theta$ theorem is used as a technical step for proving finite sample bounds on the eigenvectors (see for instance [Chatterjee, 2015, Levin and Lyzinski, 2017, Tang et al., 2013]). We will prove finite sample guarantees for the Frobenius error of our estimator, under the spectral gap assumption. In particular, we will prove that under certain Sobolev regularity assumptions the rate of convergence for the proposed Gram matrix estimator will be parametric. Hence, the results presented here not only extend the approach developed in [Araya and De Castro, 2019], but also improve the convergence rate. The proof will be mainly based on the harmonic analysis on $B^d$ and matrix concentration inequalities for the operator norm[Tropp, 2012, Vershynin, 2012, Bandeira and Van Handel, 2016].

It is worth mention that some related problems, involving the recovery of latent structures, have recently been studied in [Athreya et al., 2020], from the spectral point of view, but on the RDPG model and with distributional assumptions of the latent points and ambient spaces different from the ones we consider here.

1.1. Notation

We will use the asymptotic notation as usual. For a real function $f$, we write $f(x) = O(g(x))$, for $g$ strictly positive, if and only if there exists $C > 0$ such that $|f(x)| \leq C g(x)$ for $x$ larger that certain $x_0$. We use the symbol $\lesssim$ to denote inequality up to constants, that is $f(x) \lesssim g(x)$ if and only if there exist $C > 0$ such that $f(x) \leq C g(x)$. Similarly, $f(x) \lesssim_\alpha g(x)$ will denote
We recall the definition of the graphon degree function \cite{Lovasz2012} [Chap. 7], which can be defined on \(\mathbb{R}^d\) valued random graph \cite{Lovasz2012} [Chap. 10] is described by a two step procedure as follows: given \(F_\nu \in \mathcal{F}\) and a link function \(f : [-1, 1] \to [0, 1]\), which we assume measurable, we first sample i.i.d latent points \(\{X_i\}_{1 \leq i \leq n}\) according to \(F_\nu \in \mathcal{F}\), for some \(\nu > 0\). Then, conditional to these latent points, we sample the adjacency matrix \(A_{ij}\) such that for \(i < j\), the entries \(A_{ij}\) are independent Bernoulli variables and

\[ P(A_{ij} = 1) = f(\langle X_i, X_j \rangle) \]

The entries \(A_{ij}\) for \(i > j\) are defined by symmetry and recall that \(A_{ii} = 0\), for all \(i \in [n]\). This model contains as subclasses some classic random graphs models such as the Erdős-Rényi model, where \(f(t) = p\) for \(p \in [0, 1]\), threshold or proximity graphon \(f(t) = 1_{t \geq \tau}\) for \(\tau \geq 0\) and the random dot product graph for \(f(t) = \frac{1}{2}(1 - t)\).

### 2.2. The degree function

We recall the definition of the graphon degree function \cite{Lovasz2012} [Chap. 7], which can be seen as analogous to the normalized degree of a node on a finite graph. Let \(W\) be a graphon defined on \(\Omega\) with measure \(\mu\), the (normalized) degree function is defined as follows

\[ d_W(x) := \int_{\Omega} W(x, y) d\mu(y) \]

In the case of the Erdős-Rényi model, that is when \(W(x, y) = p\), for some \(p \in [0, 1]\), we have that \(d_W(x) = p\), \(\forall x \in \Omega\) (which is valid for any measurable space \((\Omega, \mu))\). In the case of a
graphon of the form $W(x, y) = f((x, y))$ defined on $\Omega = S^{d-1}$ with $\mu = \sigma$, where $\sigma$ is the uniform measure on $S^{d-1}$, we have that $d_W(x)$ is also constant (this follows by a simple change of variables). When $W(x, y) = f((x, y))$ and $\Omega = B^d$ and $\mu = F_\nu$, for some $\nu > -\frac{1}{2}$, we see that the $d_W(x) = d_W(x')$ for all $x, x' \in B^d$, such that $\|x\| = \|x'\|$. Take for instance the threshold function $W_\theta(x, y) = 1_{(x, y) \geq \tau}$, for some $\mu = F_\nu$. Then we have for the degree function

$$d_W(x) = \int_{B^d} 1_{(x, y) \geq \tau} dF_\nu(x)$$

$$= F_\nu \left( \operatorname{Sc}(x, 1 - \frac{\tau}{\|x\|}) \right)$$

where $\operatorname{Sc}(x, h)$ represents the spherical cap on $x/\|x\|$ with heigh $h$, that is

$$\operatorname{Sc}(x, h) := \{ y \in B^d : \langle y, x/\|x\| \rangle \geq 1 - h \}$$

Fix $X_i \in B^d$, then the probability that $X_j$ is connected to $X_i$ for $j \neq i$ is

$$P(A_{ij} = 1) = F_\nu \left( \operatorname{Sc}(X_i, 1 - \frac{\tau}{\|X_i\|/\|x\|}) \right)$$

Note that if $\|X_i\| \leq \tau$, then the spherical cap in the previous formula reduce to a point and, therefore, has measure zero. In other words, the points $X_i$ such that $\|X_i\| < \tau$ are associated with isolated nodes and the points such that $\|X_i\| = \tau$ are almost surely isolated. The degree function on a graphon can be regarded as the continuous analog to the normalized degree on a finite graph. To make this more precise, we denote $d_G(X_i) := \sum_{j \neq i} A_{ij}$ the degree of $X_i$ in the random graph. Observe that the random variable $d_G(X_i)$, conditional to $X_i$, follows a distribution $\text{Binomial}(n - 1, d_W(X_i))$, thus

$$E \left( \frac{d_G(X_i)}{n - 1} \right) = d_W(X_i) = F_\nu(\operatorname{Sc}(X_i, 1 - \tau/\|X_i\|))$$

(1)

We will use the degree of an observed graph (from the RGG on $B^d$ model) to deduce the latent norm in the following way. First, from standard concentration inequalities, we deduce for each $i$ the degree is highly concentrated around its mean $d_W(X_i)$. From the spherical symmetry of $F_\nu$, we deduce that for each $i$, the right hand side of (1) depends only on $\|X_i\|$ and from Lemma 12 we deduce the explicit form of the relation (and its inverse) that maps $\|X_i\|$ into $d_W(X_i)$. From this, we define an estimator of $\|X_i\|$ based on the degree of the node $i$ and proves its consistency (in Proposition 2).

We will also use the degree function to prove that for certain link functions, the degree sequence presents tails that decay as a power law. As pointed in [Janson, 2018, Sec.9], there is no standard definition of graph with power-law distributed degrees. While power-law for the degree are mentioned in the graphon literature, such as [Borgs et al., 2018], no precise definition is given. We will introduce the following definition

**Definition 1.** Given a graphon $W$, defined on $(\Omega, \mu)$, we will say that its degree has power law tails if there exist $0 \leq \kappa < 1$, $C > 0$ and $\theta > 0$ such that

$$\mu(\{ x \in \Omega : d_W(x) \geq h \}) \geq C(h - \kappa)^{-\theta}$$

for $\kappa < h < 1$.

---

1With some abuse of notation we use $F_\nu$ for the distribution function and the measure.
3 MAIN RESULTS

Remark 1. In the case $\kappa > 0$, it can be said that the degree is distributed according to a shifted power-law.

In Section 5 we will study the power law property for some specific graphons on $B^d$.

3. Main results

Here we gather the main results of this paper. We start by considering the case of graphs generated by the threshold graphon with parameter $\tau > 0$, that it $W_g(x, y) = \mathbb{1}_{\langle x, y \rangle \geq \tau}$. Given the observation of graph of size $n$, which we suppose generated by the $W$-random graph model with $W_g$ and latent points distributed according to some $F_\nu \in \mathcal{F}$, we want to obtain information about the latent points $X_i$. In particular, we are interested in estimating the norm of latent points $\|X_i\|_{1 \leq i \leq n}$ and the Gram matrix $G^*$, which has entries $G^*_{ij} = \langle X_i, X_j \rangle$.

It is not possible to estimate the latent norms, or any positional information for that matter, if no restriction on the measure is imposed. That is, there exist different combinations of measures in $\mathcal{F}$ and link functions that generate the same random graph model (the same distribution over finite graphs). This is given in Proposition 15, below.

We consider the threshold graphon model on $B^d$, defined by the kernel $W_g(\langle x, y \rangle) = \mathbb{1}_{\langle x, y \rangle \geq \tau}$ for a fixed $F_\nu \in \mathcal{F}$ and $\tau > 0$. We assume that a graph $G$ is observed from that model. For each node $i$ in $G$ we define $Z_i := I - \frac{1}{n-1} (2d_{G(i)}; \nu + \frac{d}{2}, \frac{1}{2})$.

For a fixed $i$, we define the following estimator of the norm $\|X_i\|$ based on $Z_i$:

$$\hat{\zeta}_i := \frac{\tau}{\sqrt{1 - Z_i}} \vee 1$$

The following proves the consistency of the estimator and it is mainly a consequence of the strong law of large numbers.

Proposition 2. For a fixed $i \in \mathbb{N}$, the random variable $\hat{\zeta}_i$ converges almost surely to $\|X_i\|$.

We now turn our attention to the problem of estimating the Gram matrix of the latent points. We will consider $G^* = \frac{1}{n}(1 - \delta_{ij})(X_i, X_j)$ the population Gram matrix (with the diagonal erased) and $\hat{G}_G := \frac{1}{2c_\nu(1 + \gamma_\nu)}U U^T$ for any $n \times d$ real matrix $U$. The reason for the chosen normalization $2c_\nu(1 + \gamma_\nu)$, comes from the harmonic analysis on $B^d$ and it will be clarified in Section 4 below. We now give a slightly informal version of our main result regarding the estimation of $G^*$, which will be stated more formally in Section 4.2.

Theorem 3 (Informal version). Let $W$ be a graphon defined by a dot product kernel on $B^d$ and measure $F_\nu \in \mathcal{F}$. If $W$ is sufficiently regular, in the Sobolev sense, and satisfy a spectral gap condition, then there exists a set of $d$ eigenvalues $\hat{v}_1, \cdots, \hat{v}_d$ of the normalized adjacency matrix of the observed graph, such that with high probability

$$\|G^* - \hat{G}\|_F \leq C(W) \frac{1}{\sqrt{n}}$$

where $\hat{G} = \hat{G}_\nu$, and $\hat{V}$ is the matrix with columns $\hat{v}_1, \hat{v}_2, \cdots, \hat{v}_d$.

In the previous theorem the constant $C(W)$ will depend on the spectral gap of $W$, which will be defined in Section 4.2.

We will now see the main elements for the proof of Theorem 3, which are related to the properties of the spectrum of the integral operator $T_W$. 
4. Graphon eigensystem and harmonic analysis on $\mathbb{B}^d$

We will extend the method developed in [Araya and De Castro, 2019] for the case of geometric graphs on the sphere. For that, it will useful to consider an orthogonal polynomials basis of $L^2(\mathbb{B}^d, F_\nu)$, with respect to the inner product given by

$$\langle f, g \rangle = a_\nu \int_{\mathbb{B}^d} f(x)g(x)dF_\nu(x)dx$$

where we recall that $dF_\nu(x) = (1 - \|x\|^2)^{\nu-\frac{d}{2}}$ and $a_\nu = 1/ \int_{\mathbb{B}^d} F_\nu(x)dx$.

We denote $\mathcal{Y}_n$ the subspace of orthogonal polynomials on $d$ variables (with respect to the inner product defined above) of degree exactly $n$. It is implicit that $\mathcal{Y}_n$ depend on $\nu$. From [Dai and Xu, 2013, p.266] we know that the space dimension is $\dim(\mathcal{Y}_n) = \binom{n+d-1}{n}$ (this actually can be seen by applying a Gram-Schmidt orthonormalization process to monomials).

There are explicit expressions (closed formulas) for the reproducing kernel $P_n^\nu(x,y)$ of each $\mathcal{Y}_n$. We recall that $P_n^\nu(x,y)$ is said to satisfy the reproducing property on $\mathcal{Y}_n$ if and only if

$$p(x) = \int_{\mathbb{B}^d} P_n^\nu(x,y)p(y)dF_\nu(y), \quad \forall p \in \mathcal{Y}_n, \forall x \in \mathbb{B}^d$$

In our context, the reproducing kernel $P_n^\nu(x,y)$ is the projector of $L^2(\mathbb{B}^d, F_\nu)$ onto $\mathcal{Y}_n$, which can be seen by writing $P_n^\nu(x,y) = \sum_{p_{k,n} \in Q_n} p_{k,n}(x)p_{k,n}(y)$, where $Q_n := \{p_{k,n}\}_{k=1}^{\dim \mathcal{Y}_n}$ is an $L^2(\mathbb{B}^d, dF_\nu)$-orthonormal basis of $\mathcal{Y}_n$. A key result is the following close form representation of the reproducing kernel [Dai and Xu, 2013, Cor.11.1.8] (see also [Xu, 2001, Eq.2.2]), for $\nu > 0$

$$P_n^\nu(x,y) = c_\nu \frac{n + \gamma_\nu}{\gamma_\nu} \int_{-1}^1 G_n^{\gamma_\nu}(\langle x, y \rangle) + \sqrt{1 - \|x\|^2}\sqrt{1 - \|y\|^2}t(1 - t^2)^{\nu-1}dt$$

(2)

where $\gamma_\nu := \nu + \frac{d-1}{2}$, $c_\nu = \frac{\Gamma(\nu+1/2)}{\sqrt{\pi}F(\nu)}$ and $G_n^{\gamma_\nu} (\cdot)$ is the Gegenbauer polynomial of degree $n$ with weight $\gamma_\nu$.

It is well known that $\{G_n^{\gamma_\nu} (\cdot)\}_{n \geq 0}$ forms a basis for $L^2(\{-1,1\}, \gamma_\nu)$ [Szego, 1939]. In addition, each $p_k \in \mathcal{Y}_n$ is an eigenfunction of the following $L^2(\mathbb{B}^d, dF_\nu)$ integral operator

$$T_f g(x) = \int_{\mathbb{B}^d} f(\langle x, y \rangle)g(y)dF_\nu(y)dy$$

for any $f \in L^1([-1,1], \gamma_\nu)$ (which is automatic in our case, given that $f$ is bounded). The previous statement is a consequence of the Funk-Hecke formula in this context [Dai and Xu, 2013, Thm.11.1.9], which also give us a formula for the eigenvalue associated to each $p_k \in \mathcal{Y}_n$ is

$$\lambda_n^\nu(f) = c_{\gamma_\nu} \int_{-1}^1 f(t) \frac{G_n^{\gamma_\nu}(t)}{G_n^{\gamma_\nu}(1)}(1 - t^2)^{\gamma_\nu-1/2}dt$$

(3)

and $c_{\gamma_\nu}$ is such that $\lambda_0^\nu(1) = 1$. Notice that for each $n \in \mathbb{N}$ we have the following decomposition of the reproducing kernel of $\mathcal{Y}_n$ in terms of the basis elements $p_{k,n}$

$$P_n^\nu(x,y) = \sum_{p_k \in Q_n} p_{k,n}(x)p_{k,n}(y)$$

(4)
hence for a given $f(\langle x, y \rangle)$ the following decomposition holds, in virtue of the spectral theorem for compact operators
\[ f(\langle x, y \rangle) = \sum_{n \in \mathbb{N}} \lambda_n^*(f) P_n^\nu(x, y) \] (5)

We note that previous implies that $\lambda_n^*(f)$ is an eigenvalue associated to every $p_{k,n} \in Q_n$, which means that it has multiplicity $\dim(\mathcal{Y}_n)$. We come back to formula (2), which in the linear case gives a representation of the reproducing kernel $P_n^\nu$ in terms of the inner product $\langle x, y \rangle$. Given that $G_1^1(t) = 2\gamma t$, we have
\[
P_n^\nu(x, y) = 2c_\nu \frac{1 + \gamma_\nu}{\gamma_\nu} \int_{-1}^1 (\langle x, y \rangle + \sqrt{1 - \|x\|^2} \sqrt{1 - \|y\|^2} t)(1 - t^2)^{\nu - 1} dt
\]
\[= 2c_\nu(1 + \gamma_\nu)\langle x, y \rangle \] (6)

where $c_\nu = c_\nu \int_{-1}^1 (1 - t^2)^{\nu - 1} dt$. In the last step we used that $\int_{-1}^1 t(1 - t^2)^{\nu - 1} dt = 0$ given the parity of the function inside the integral. From formula (4) we deduce that
\[
\frac{1}{2c_\nu(1 + \gamma_\nu)} \sum_{p_{k,n} \in Q_n} p_{k,n}(x)p_{k,n}(y) = \langle x, y \rangle \] (7)

The previous relation has its analogous in the case of dot product kernels on $S^{d-1}$, see [Araya and De Castro, 2019]. We can read formula (7) as a presentation for the Gram matrix of the latent points in terms of the elements of the orthonormal basis of $\mathcal{Y}_1$ (which has exactly $d$ elements), which are eigenfunctions of $T_f$. We will see that, in the case of the eigenvalues of the matrix $T_n$, a finite sample analog holds.

Recalling that a $L^2$ basis of eigenfunctions of $T_W$ is given by $\bigcup_{n \geq 0} \bigcup_{p_{k,n} \in Q_n} p_{k,n}$, the following estimate will useful for proving the concentration of the eigenvectors of the adjacency matrix of the observed graph with respect to the eigenfunctions of $T_W$.

**Lemma 4.** We have for $\{p_{k,n}\}_{k=1}^{\dim(\mathcal{Y}_n)} \in \mathcal{Y}_n$
\[
\|p_{k,n}\|_\infty \leq n^{\nu + \frac{d-1}{2}} \quad \text{for } 1 \leq k \leq \dim(\mathcal{Y}_n)
\]
\[
\left\| \sum_{k=1}^{\dim(\mathcal{Y}_n)} p_{k,n}^2 \right\|_\infty \leq n^{2\nu + d - 1}
\]

**Remark 2.** Notice that when $\nu = 0$, the space of orthogonal polynomials $\mathcal{Y}_n$ coincide with $\mathcal{H}_n^d$, the space of classic spherical harmonics of degree $n$ in $S^{d-1}$. Replacing $\nu = 0$ in Lemma 4, we recover the classic estimate $\|p_{k,n}^2\|_\infty \leq n^{d-1} = \dim(\mathcal{H}_n^d)$.

We will use weighted Sobolev spaces to define regularity, which is similar to spherical context treated in [De Castro et al., 2020],[Araya, 2020].

**Definition 5.** We will say that $f \in S^p_{\nu_n}([-1, 1])$ or, equivalently, that $f$ is Sobolev regular with parameter $p$, if and only if the eigenvalues $\lambda_n^*$ (given by (3)) satisfy $\sum_{n \geq 0} |\lambda_n^*|^2 d_n(1 + \nu_n^p) < \infty$, where $\nu_n = n(n + 2\nu + d - 1)$. 


4.1. Eigenvalues and eigenvectors

We consider the integral operator $T_W : L^2(B^d, F_\nu) \to L^2(B^d, F_\nu)$

\[ T_W g(x) = \int_{B^d} f(\langle x, y \rangle) g(y) dF_\nu(y) \]

and the $n \times n$ symmetric matrices

\[ T_n = \frac{1}{n} (1 - \delta_{ij}) f(\langle X_i, X_j \rangle) \]

\[ \hat{T}_n = \frac{1}{n} A_{ij} \]

where $A_{ij}$ is the adjacency matrix defined in Section 2. The following two results are key steps to prove that the spectrum of $\hat{T}_n$ is close to the spectrum of $T_W$ with high probability

- We will use Bandeira-Van Handel result [Bandeira and Van Handel, 2016, Cor.3.12], which provides a sharp concentration inequality for the operator norm of random matrices with independent entries. This allow us to say that $\lambda(\hat{T}_n)$ is close to $\lambda(T_n)$ for $n$ large. Framed in our context, this results gives (see Thm. 17)

\[ \|\hat{T}_n - T_n\|_{op} \lesssim \alpha \frac{1}{\sqrt{n}} \]

with probability larger than $1 - \alpha$.

- In order to prove that $\lambda(T_n)$ is close to $\lambda(T_W)$ in the finite sample sense, we will improve upon the concentration result [De Castro et al., 2020, Prop.4], which says that, with high probability, $\delta_2(\lambda(T_n), \lambda(T_W))$ decrease as $n$ grows, with a nonparametric rate depending on Sobolev-type regularity conditions (analogous to Def.5). More formally, in Proposition 19 we prove that

\[ \delta_2(\lambda(T_n), \lambda(T_W)) = O_\alpha\left(\frac{1}{\sqrt{n}}\right) \]

provided that $W$ is Sobolev regular with parameter $p > 2\nu - 1 + \frac{5d}{2}$.

4.2. Eigengap condition

In context of Gram matrix estimation, will assume the following eigengap condition. Given a geometric graphon $W$, we define the spectral gap of $W$ relative to the eigenvalue $\lambda^*_1$ by

\[ \Delta^*(W) := \min_{j \neq 1} |\lambda^*_j - \lambda^*_1| \]

which quantifies the distance between the eigenvalue $\lambda^*_1$ and the rest of the spectrum. We will drop the dependency on $W$ when is clear from the context. We will ask for $\Delta^*$ to be strictly positive, which will allow us to identify a cluster of eigenvalues of $\hat{T}_n$, that is a set of eigenvalues which are close to $\lambda^*_1$ and which are sufficiently isolated from the rest of the spectrum. In this case, the size of the cluster associated with $\lambda^*_1$ is exactly $\dim(\mathcal{Y}_1)$ (which is equal to $d$).

We will define the following event

\[ \mathcal{E} := \left\{ \delta_2(\lambda(T_n), \lambda(T_W)) \vee \frac{2^9 \sqrt{d}}{\Delta^*} \|T_n - \hat{T}_n\|_{op} \leq \frac{\Delta^*}{4} \right\}, \]

for which we prove the following
Lemma 6. Assume that $\Delta^* > 0$, then there exists $n_0 \in \mathbb{N}$ such that for $n \geq n_0$ and for $\alpha \in (0,1)$ we have with probability larger than $1 - \alpha$

$$P(\mathcal{E}) \geq 1 - \frac{\alpha}{2}$$

Proposition 7. On the event $\mathcal{E}$, there exists one and only one set, consisting of $d$ eigenvalues of $\hat{T}_n$, whose diameter is smaller than $\Delta^*/2$ and whose distance to the rest of the spectrum of $\hat{T}_n$ is at least $\Delta^*/2$.

We now give the main result of gram matrix estimation, which is a more precise version of Thm.3.

Theorem 8. Let $W$ be a graphon defined by a dot product kernel on $\mathbb{B}^d$ and measure $F_{\nu} \in \mathcal{F}$. If $W \in S_{2p \nu}([-1,1])$ for $p > 2\nu - 1 + \frac{5d}{2}$ and we assume that $\Delta^*(W) > 0$, then there exists a set of $d$ eigenvalues $\hat{v}_1, \cdots, \hat{v}_d$ of $\hat{T}_n$, such that with probability larger than $1 - \alpha$ we have

$$\|G^* - \hat{G}\|_F \lesssim \Delta^*(W)^{-1} \frac{1}{\sqrt{n}}$$

where $\hat{G} = \hat{G}_{\hat{V}}$ and $\hat{V}$ is the matrix with columns $\hat{v}_1, \hat{v}_2, \cdots, \hat{v}_d$.

Remark 3. The condition $p > 2\nu - 1 + \frac{5d}{2}$ is mainly technical, as it is a sufficient condition for having a parametric rate for $\delta(\lambda(T_n), \lambda(T_W))$, see Proposition 19. The dependency on alpha in (8) is through a multiplicative constant of the form $\log 1/\alpha$.

5. The sough after power law distribution

As we already announced in the introduction, the RGG model on the ball is more flexible that the one on the sphere, in terms of observed degree distribution, when we restrict ourselves to spherically symmetric distributions (which is desirable in light of the previous chapter). Given that the heterogeneity in the degree sequence is a characteristic observed in many real world networks [Barrat et al., 2004], having a more flexible model at hand would be useful for modeling purposes. From (1) we see that the possible values for $d_W(x)$ for a threshold graphon $W$ are of the form $F_{\nu}(Sc(N, 1 - \frac{1}{\|x\|\nu^2}))$. However useful the previous characterization might be, it has the problem of not being very explicit and as it is written do not match any of the typical degree distributions that are frequent in the network literature. In particular, many real networks exhibit a power law degree distribution [Clauset et al., 2009, Mitzenmacher, 2003], meaning that the number of nodes with (unnormalized) degree $k$ is proportional to $k^{-\gamma}$ with $\gamma > 0$. This opens the question: is the power between the possible degree distributions in the RGG model on the ball? or at least, is it possible to have an approximative version of it?

We will consider the following RGG on $\mathbb{B}^d$, defined by the connection function:

$$f(t) = \begin{cases} \frac{\alpha}{|t|^\gamma} \wedge 1 & \text{if } t \neq 0 \\ 1 & \text{if } t = 0 \end{cases}$$

where $\alpha \in (0,1)$ is a “resolution” parameter. For the latter we mean that if $x \in \mathbb{B}^d(0, \sqrt{\alpha})$ then for all $y \in \mathbb{B}^d$ we have $f((x,y)) = 1$. That is, any point located in the ball centered in 0 with radius $\alpha$ will connect with every other point in $\mathbb{B}^d$. This is the inverse of what happens in the
threshold graphon. In terms of the degree function, this means that \( d_f(x) = 1 \) for \( \|x\|^2 \leq \alpha \).

By definition we have

\[
d_f(x) = \int_{\mathbb{B}^d} \frac{\alpha}{\|x,y\|^2} \wedge 1 \, dF_\nu(y)
\]

By rotational symmetry (we can think of \( x \) being \( x = (x_1, 0, \cdots, 0) = x_1 N \)) we have

\[
d_f(x) = d_f(x_1 N)
= \int_{\mathbb{B}^d} \frac{\alpha}{\|x,y\|^2} \wedge 1 \, dF_\nu(y)
= \int_{\mathbb{B}^d \setminus \mathbb{B}^d(0,\sqrt{\alpha})} \frac{\alpha}{\|x_1\|^2} \wedge 1 \, dF_\nu(y) + \int_{\mathbb{B}^d(0,\sqrt{\alpha})} \, dF_\nu(y)
\]

Given that the summand \( \int_{\mathbb{B}^d(0,\sqrt{\alpha})} dF_\nu(y) \) is common to every \( x \in \mathbb{B}^d \) we will subtract it.

Intuitively speaking, there will be nodes that will be connected will almost every node in the graph, which increase the minimum degree. Since we already know that the nodes such that \( \|x\|^2 \leq \alpha \) are connected with every other node, we concentrate in the case \( \|x\|^2 > \alpha \). This motivates the definition

\[
\tilde{d}_f(x) := \frac{\int_{\mathbb{B}^d \setminus \mathbb{B}^d(0,\sqrt{\alpha})} \frac{\alpha}{\|x_1\|^2} \, dF_\nu(y)}{\int_{\mathbb{B}^d \setminus \mathbb{B}^d(0,\sqrt{\alpha})} \frac{\alpha}{\|x_1\|^2} \wedge 1 \, dF_\nu(y)}
\]

for \( \|x\|^2 > \alpha \). Let take \( k, n' \in \mathbb{N} \) such that \( k < n' \leq n \). By definition

\[
\tilde{d}_f(\sqrt{\frac{n'\alpha}{k}N}) = \frac{\int_{\mathbb{B}^d \setminus \mathbb{B}^d(0,\sqrt{\alpha})} \frac{k}{\|x_1\|^2} \wedge 1 \, dF_\nu(y)}{\int_{\mathbb{B}^d \setminus \mathbb{B}^d(0,\sqrt{\alpha})} \frac{1}{\|x_1\|^2} \, dF_\nu(y)}
= \frac{k}{n'}
\]

Thus, given that \( X \) is distributed according to \( F_\nu \)

\[
\mathbb{P}\left(\frac{k - 1}{n'} \leq d_f(X) \leq \frac{k}{n'}\right) = \mathbb{P}(X \in \mathbb{B}^d(0, \sqrt{n'\alpha/k - 1}) \setminus \mathbb{B}^d(0, 1 / k))
= F_\nu\left(\mathbb{B}^d(0, \sqrt{1/k - 1}) \setminus \mathbb{B}^d(0, \sqrt{1/k})\right)
= I_{n'/\alpha}(\frac{d}{2}, \nu + \frac{1}{2}) - I_{n'/\alpha}(\frac{d}{2}, \nu + 1/2)
\]

Since \( I_\nu(\cdot, \cdot) \) is an increasing function we see that

\[
\Delta_k \frac{d}{dx} I\left(\frac{n'\alpha}{k - 1} \right) \leq I\left(\frac{n'\alpha}{k - 1} \right) - I\left(\frac{n'\alpha}{k} \right) \leq \Delta_k \frac{d}{dx} I\left(\frac{n'\alpha}{k} \right)
\]

where \( I(x) = I_\nu(\frac{d}{2}, \nu + \frac{1}{2}) \) and \( \Delta_k = \frac{n'\alpha}{k - 1} - \frac{n'\alpha}{k} \). It follows from the definition that \( \frac{d}{dx} I_\nu(a, b) = \frac{1}{B(a, b)} x^{a-1} (1 - x)^{b-1} \) and consequently

\[
c_{\nu,\alpha}(\frac{k - 1}{n'})^{-\frac{d}{2} - 1} \leq I\left(\frac{1}{k - 1} \right) - I\left(\frac{1}{k} \right) \leq c_{\nu,\alpha}(\frac{k}{n'})^{-\frac{d}{2} - 1}
\]
where \( c_{\nu,\alpha}, C_{\nu,\alpha} \) are constants that depend on \( \nu \) and \( \alpha \). Thus picking \( d = 3 \), for example, we have that
\[
P\left( \frac{k - 1}{n'} \leq \tilde{d}_f(X) \leq \frac{k}{n'} \right) \propto (k/n')^{-2.5}
\]
which can see as a similar distribution to a power law\(^2\). To make this point clearer, take for example \( n' \) of order \( \mathcal{O}(n) \).

The previous can be interpreted as the proportion of nodes of degree \( k \), for \( k \) large, follows a power law function, after shifting. The exponent \(-2.5\) has been frequently reported in the literature for real networks \[\text{Clauset et al., 2009}\]. We see that that changing the exponent \( \alpha \) in the definition of \( f(t) \) and the changing the dimension of the sphere, we can fine-tune the power law exponent parameter.

\[\sqrt{\nu} \sim k/n' \]

Fig 1. A point \( x \) inside the annulus between the circles \( \sqrt{n'_{\nu}/k^{1/\nu}} \) and \( \sqrt{n'_{\nu}/k} \) will have degree function satisfying \( \tilde{d}_f(x) \sim k/n' \). The fraction of points with degree \( k/n' \) is the measure of the annulus.

The following proposition proves that \( f \) has power law tails also in the sense of Definition 1

**Proposition 9.** There exist a constant \( C > 0 \) such that the degree function \( d_f(x) \) satisfies
\[
F_\nu(\{ x \in B^d : d_f(x) \geq h \}) \leq C(h - \kappa)^{-\theta}
\]
for \( \theta = 1.5, \kappa = \int_{B^d(0,\sqrt{\nu})} dF_\nu(y) \).

The convergence of the cumulative distribution of the degrees is proven in [Delmas et al., 2018] and [Borgs et al., 2018]. In [Borgs et al., 2018], the authors prove the convergence of \( |\{i \in [n] : d_{G_n}(X_i) > h\}| \) towards \( \mu(\{y : d_W(y) > h}\}) \), where \( \mu \) is the distributions of the points \( X_i \) and \( \lambda > 0 \) is a point of continuity of \( \lambda \to \mu(\{d_W(y) > h\}) \).

In [Delmas et al., 2018], a graphon \( W \) in \([0,1]\) is considered and the convergence of \( \Pi(y) := \frac{1}{n} \sum_{i=1}^n 1_{d_G(X_i) < nd_W(y)} \) toward \( y \), almost surely, is proved. They also provide a CLT type result for this convergence.

From the previous we can deduce that if \( \{G_n\}_{n \geq 1} \) is a sequence of graphs, obtained by sampling the graphon \( W(x,y) = f((x,y)) \) by the \( W \)-random graph model, as described in Sec. 2, then there exists a \( n_0 \in \mathbb{N} \) such that all the graphs \( \{G_n\}_{n \geq n_0} \) will have a (discrete) power law type distribution. Indeed, this is consequence of the characterization of \( F_\nu(\{ x \in B^d : d_W(x) \geq h\}) \) given in this section, the fact that the sequence \( G_n \) will converge to \( W \) in the cut distance [Lovasz, 2012] and [Borgs et al., 2018, Prop.21].

\(^2\)A random graph model has power law if the number of nodes with (unnormalized) degree \( k \) is proportional to \( k^{-\gamma} \) with \( \gamma > 0 \).
6. Numerical Experiments

We run simulation for different RGG models on $\mathbb{B}^d$ and compute the estimators analyzed throughout this paper (for the norm and the Gram matrix) to see how they perform empirically. In the case of the Gram matrix estimation, we run the algorithm HEiC described in [Araya and De Castro, 2019, Sec.3], changing the normalization constants as described in Section 4.

We start by considering simulations for the link function that gives a power-law type degree distribution, considered in Section 5.

6.1. Power law type graphon

We consider the link function

$$f(t) = \begin{cases} \frac{\alpha}{|t|^{2}} \wedge 1 & \text{if } t \neq 0 \\ 1 & \text{if } t = 0, \end{cases}$$

to illustrate empirically the behavior of the degree profile under this model. In Figure 2 we show a single simulation of the graph of size 3000 with connection function $f(\cdot)$ and parameter $\alpha = 1/1000$, under the measure $F_1/2(\cdot)$, which we recall is the uniform measure in $\mathbb{B}^d$. We observe the presence of nodes with very high degree (or large hubs) which is often observed in real world networks and scale-free networks. We include a log – log plot for the histogram for the nodes with degrees over 300, to better observe the exponential decay. The resulting shape, first close to a line and then oscillatory (Figure 2(right)), has been reported in real world networks, where it is suggested as evidence for a power law distribution of the degrees [Clauset et al., 2009].

![Fig 2](image)

Fig 2. From left to right: we plot the function $f(\cdot)$ for $\alpha = 1/1000$. In the center, we show the histogram for this $f$. In the right, we show a log – log plot of the same histogram, but only for the values with degree larger than 300.

We repeat this exercise in Figure 3, for different values of $\alpha$ which produces changes in the distribution. We opt to include the log – log plot for nodes with degree larger than 300 for comparison purposes. This shows the shifted power law shape of the degree distribution. More node connectivity can also be achieved by changing the measure under which we simulate. We show one example in the image at the bottom in Figure 3, which was generated with $F_{3/2}$. Indeed, a measure that allocates more mass at the center, will have larger connectivity within this model. This serves to illustrate the flexibility and expressiveness of this model.
6 NUMERICAL EXPERIMENTS

Fig 3. Analogous to Figure 2, using different values of $\alpha$. In the log–log plot we show the values with degree larger than 300. The plot at the bottom was done with a different measure $F_{3/2}$.

6.2. Latent norm recovery

We study the empirical performance of the estimator $\tilde{\zeta}_i$, for each $1 \leq i \leq n$, for which we proved almost sure converge to the latent norm on the threshold RGG model. We compute the estimator for each node in the graph, following measure of error for each sample

$$E_{\text{norm}} = \frac{1}{\sum_{i=1}^{n} \mathbb{1}_{\|X_i\| \geq \tau}} \left( \sum_{i=1}^{n} (\zeta_i - \|X_i\|)^2 \mathbb{1}_{\|X_i\| \geq \tau} \right)^{1/2}$$

We discard the points with norm larger than $\tau$, because as discussed in Section 3 the adjacency matrix of the graph carries no information about the norm of those points, other than being smaller than $\tau$. In Figure 4(left) we plot the mean square error $E_{\text{norm}}$ in logarithmic scale for a threshold $\tau = 0.1$. For each sample size, we simulate 25 graphs on the ball with dimension $d = 3$, and uniform measure $F_{1/2}$, and compute the mean of the errors $E_{\text{norm}}$. The form in which the error decrease suggest a parametric rate of convergence, which we plot in a red line for reference. However, note that the fact the estimator is based in a complicated nonlinear function, as it is

$$t \rightarrow \frac{\tau}{\sqrt{1 - I^{-1}(t; \frac{d}{2} + \nu, \frac{1}{2})}}$$

makes that this rate is non-uniform across the nodes. Indeed, given the shape of the graph of $I^{-1}(t; \frac{d}{2} + \nu, \frac{1}{2})$ it is not hard to see that points in with higher norm (closer to 1) will converge...
slower. This indeed what we observe in the experiments as shown in Figure 4(right), where we plot the sequence of ordered norms in red and the sequence of ordered $\zeta_i$’s for different values of the sample size ($n = 100$). Notice that it takes much more samples to see a convergence when the norm of the node is close to 1.

We observe that for values of $\tau$ closer to 0, the convergence is indeed slower. In Figure 4 we plot the mean of $\log(E_{\text{norm}})$ over 25 sampled graphs, for a threshold $\tau = 0.01$ with dimension parameter $d = 3$ and the measure $F_{1/2}$. We observe that it takes much more samples to converge. Even if the decrease of the errors suggest a similar parametric rate in the case of the model with smaller $\tau$, the constant (intercept) is larger, which means that the error is always larger than in the previous case. This should not be surprising given that we know that in the model with $\tau = 0$ we cannot infer the norm from the samples (as the model is equivalent to the threshold graphon on the sphere). Approaching to $\tau = 0$ will render the problem harder, in the sample complexity sense.

**Fig 4.** In the left we show the mean of $\log(E_{\text{norm}})$ over 25 graphs, for the recovery of the norm on the threshold graph with $\tau = 0.1$ and parameter $d = 3$ and measure $F_{1/2}$. We add the upper in red for reference. In the right we plot the sequence of ordered values for $\|X_i\|$ in increasing order together with the sorted sequence of estimated values $\zeta_i$.

**Fig 5.** This is analogous to Figure 4, with $\tau = 0.01$ and maintaining the rest of parameters.

To see empirically the effect of changing the parameter $\nu$ in the estimation of the norm, in Figure 6 we plot the mean of the error $\log(E_{\text{norm}})$ across 25 samples for the threshold graphon
with $\tau = 0.1$ and $d = 3$. We see that a larger $\nu$ gives lower error, this is explained by the fact the larger the $\nu$, the more concentrated the sampled nodes are close to the center of the ball. We added, for reference, the plot of the theoretical density of the (squared) norm of the latent points (a Beta distribution by Lemma 10) in Figure 6 (right).

![Graph showing error for threshold graphon with $\tau = 0.1$, $d = 3$, and changing measure $F_\nu$.](image)

**Fig 6.** Similar to Figure 5. We plot the error for the threshold graphon with $\tau = 0.1$, $d = 3$ and changing the measure $F_\nu$. In the right we plot the pdf of $\text{Beta}(\frac{d}{2}, \nu + \frac{1}{2})$ which is the distribution of the norm of the nodes.

### 6.3. Gram matrix estimation

We report the empirical performance of the algorithm HEiC, described in [Araya and De Castro, 2019] applied in the context of RGG in $\mathbb{R}^d$. Similar to the spherical case, we will mainly measure the mean error

$$ME_n = \|G^* - G\|_F$$

We first consider the threshold graphon with parameter $\tau = 0.1$ in dimension $d = 3$. We sample 25 graphs using this model and run each time the algorithm HEiC. In Figure 7(left) we show a boxplot for $\log(ME_n)$ for different sample sizes. In Figure 7(right) we show the $\log(ME_n)$ error for different values of $n$ in the case of the logistic graphon $f(t) = \frac{1}{1 + e^{rt}}$ for different values of $r$. The curves in the plot were obtaining by averaging across 25 samples for each value of $n$. We observe that for $r = -0.1$ the error does not decrease with the sample size, which is to be expected as the logistic function for that value of $r$ is close to a constant function. In our parametrization of the problem, this translate into a close to 0 spectral gap as the Figure 8 illustrates. Indeed, we plot the first 10 eigenvalues, for this case the cluster of eigenvalues associated to $\lambda^*_1$ is a subset of the first 10 eigenvalues. We see that as $r$ is closer to zero, the spectral gap decrease and the number of samples required to have a decreasing error increase.

![Graph showing first 10 eigenvalues.](image)

Note that Theorem 8 do not give information about the diagonal of the Gram matrix, which corresponds to the square of the norms of the nodes $X_i$. Our measure of error $ME_n$ do not take them into consideration. In the case of the threshold graphon we can use the estimator $\zeta_i$ to compute the means. We observed empirically that the algorithm works better when the rows matrix of eigenvectors $\hat{V}$, which has columns $v_1, \cdots, v_d$ which are the output of the algorithm HEiC, are normalized to match the mean of the true means $\zeta_i$. This is not an ideal situation from the practical point of view, given that the norms are usually non available. In the case of the threshold graphon we can use the estimated norms in this extra normalization step. While this gives reasonable results in practice, a thorough theoretical study is lacking at this moment and will be left for future work.
Fig 7. In the left a boxplot for $\log(M_{E_n})$, for different values of the sample size in threshold graphon with $d = 3, \tau = 0.1$ and $F_{1/2}$. In the right we plot the mean error for the logistic graphon with different values of the parameter $r$.

Fig 8. We plot the first 10 eigenvalues for the logistic graphon for different values of the parameter $r$. We include the spectral gap in each case. In all the examples we used a parameter $n = 1000$.

**Remark 4.** The time complexity (or running time) of the latent distance recovery algorithm does not increase, in comparison with the spherical case, and it is roughly $O(n^3 + n)$. In the case of the computation of the estimators $\zeta_i$ we need to compute the degrees, which corresponds to compute the sum of all rows, which is roughly $O(n^2)$.

7. Conclusions and future work

We studied the problem of estimating the norm and the Gram matrix for the latent points of graphs generated by the RGG model on $B^d$. Extending the approach of Proposition 2 to known (given) link functions other than the threshold function is possible, because in that case we will have an expression analog to (1). On the other hand, we expect that the use of global information, in conjunction with the degree function, would help us to find simpler estimators which are more prone to be studied, in the finite sample setting, with the standard concentration tools.

The problem of estimating $\tau$, under the model with threshold link function $W_g(x,y) = 1_{(x,y) \geq \tau}$ (for a given $F_\nu$), is also of interest. This problem has been studied in the model with $\Omega = [-1,1]^2$ and link function $1_{\|x-y\| \leq r(n)}$ in [Diaz et al., 2018], where the
uniform distribution is considered, but the model allows for sparser graphs. They propose an estimator based on the explicit formula for expectation of the number of edges. In the context of the model we presented here, we believe that simpler estimators are possible, given the fact isolated nodes give information about $\mathbb{E} \nu(B_d(0, \tau))$. The main difficulty will be to estimate, with high probability, the number of isolated nodes whose associated points are outside $B_d(0, \tau)$. Constructing such estimators is left for future work.

Another interesting problem will be the estimation of the parameter $\alpha$ for the link function

$$f(t) = \begin{cases} \frac{\alpha}{|t|^2} \wedge 1 & \text{if } t \neq 0 \\ 1 & \text{if } t = 0, \end{cases}$$

which present a power-law type distribution for the degree. Finding a larger class of link functions satisfying this property, and a proper description of this class, will also be of interests. Eventually, the problem could be framed as a non-parametric graphon estimation and, given the Fourier-Gegenbauer decomposition analyzed in Sec. 4, it will be possible to use an spectral approach similar to the one developed in [De Castro et al., 2020].

References

[Araya, 2020] Araya, E. (2020). Relative concentration bound for the spectrum of kernel matrices. ArXiv.

[Araya and De Castro, 2019] Araya, E. and De Castro, Y. (2019). Latent distance estimation for random geometric graph. Advances in Neural Information Processing Systems, pages 8721–8731.

[Athreya et al., 2018] Athreya, A., Fishkind, D., Tang, M., Priebe, C., Park, Y., Vogelstein, J., Levin, K., Lyzinski, V., Qin, Y., and Sussman, D. (2018). Statistical inference on random dot product graphs: a survey. Journal of Machine Learning Research, 18:1–92.

[Athreya et al., 2020] Athreya, A., Tang, M., Park, Y., and Priebe, C. (2020). On estimation and inference in latent structure random graphs. arXiv:1806.01401.

[Bandeira and Van Handel, 2016] Bandeira, A. and Van Handel, R. (2016). Sharp nonasymptotic bounds on the norm of random matrices with independent entries. Annals of Probability, 44(4):2479–2506.

[Barrat et al., 2004] Barrat, A., Barthélemy, R., Pastor-Satorras, R., and Vespignani, A. (2004). The architecture of complex weighted networks. PNAS, 101(11):3747–3752.

[Borgs et al., 2018] Borgs, C., Chayes, J., Cohn, H., and Holden, N. (2018). Sparse exchangeable graphs and their limits via graphon processes. Journal of Machine Learning Research, 18:1–71.

[Borgs et al., 2008a] Borgs, C., Chayes, J., Lovasz, L., Sos, V., and Vesztergombi, K. (2008a). Convergent sequences of dense graphs i: subgraph frequencies, metric properties, and testing. Adv. Math, 219:1802–1852.

[Borgs et al., 2008b] Borgs, C., Chayes, J., Lovasz, L., Sos, V., and Vesztergombi, K. (2008b). Convergent sequences of dense graphs i. subgraph frequencies, metric properties and testing. Adv. Math, 219(6):1801–1851.

[Bubeck et al., 2016] Bubeck, S., Ding, J., Eldan, R., and Rácz, M. (2016). Testing for high dimensional geometry in random graphs. Random Structures and Algorithms, 49:503–532.

[Chatterjee, 2015] Chatterjee, S. (2015). Matrix estimation by universal singular value thresholding. Annals of Statistics, 43(1):177–214.
[Clauset et al., 2009] Clauset, A., Shalizi, C., and Newman, M. (2009). Power law distributions in empirical data. *SIAM review*, 51(4):661–703.

[Dai and Xu, 2013] Dai, F. and Xu, Y. (2013). *Approximation theory and harmonic Analysis on spheres and balls*. Springer Verlag Monographs in Mathematics.

[De Castro et al., 2020] De Castro, Y., Lacour, C., and Pham Ngoc, T. (2020). Adaptive estimation of nonparametric geometric graphs. *Mathematical Statistics and Learning*.

[Delmas et al., 2018] Delmas, J., Dhersin, J., and Sciauveau, M. (2018). Asymptotic for the cumulative distribution function of the degrees and homomorphism densities for random graphs sampled from a graphon. *arXiv:1807.09989*.

[Diaz et al., 2018] Diaz, J., McDiarmid, C., and Mitsche, D. (2018). Learning random points from geometric graphs or orderings. *arXiv:1804.10611*.

[Janson, 2018] Janson, S. (2018). On edge exchangeable random graphs. *J. Stat. Phys.*, 173:448:484.

[Kelker, 1970] Kelker, D. (1970). Theory of spherical distributions and a location-scale parameter generalization. *The indian journal of statistics, Serie A.*, 32(4):419–430.

[Levin and Lyzinski, 2017] Levin, K. and Lyzinski, V. (2017). Laplacian eigenmaps from sparse, noisy similarity measurements. *IEEE Transactions on Signal Processing*, 65:1998–2003.

[Lovasz, 2012] Lovasz, L. (2012). *Large networks and graph limits*. Colloquium Publications (AMS).

[Lovász and Szegedy, 2006] Lovász, L. and Szegedy, B. (2006). Limits of dense graph sequences. *J. Combin. Theory. Ser B*, 96(6):197–215.

[Mitzenmacher, 2003] Mitzenmacher, M. (2003). A brief history of generative models for power law and lognormal distributions. *Internet Math.*, 1(2):226–251.

[Sussman et al., 2014] Sussman, D., Tang, M., and Priebe, C. (2014). Consistent latent position estimation and vertex classification for random dot product graphs. *IEEE transactions on Pattern Analysis and Machine Intelligence*, 36:48–57.

[Szego, 1939] Szego, G. (1939). *Orthogonal polynomials*. Colloquium Publications (AMS).

[Tang et al., 2013] Tang, M., Sussman, D., and Priebe, C. (2013). Universally consistent vertex classification for latent position graphs. *Annals of Statistics*, 41:1406–1430.

[Tropp, 2012] Tropp, J. (2012). User-friendly tail bounds for sums of random matrices. *Foundations of Computational Mathematics*, 12(4):389–434.

[Vershynin, 2012] Vershynin, R. (2012). How close is the sample covariance matrix to the actual covariance matrix? *Journal of Theoretical Probability*, 25:655–686.

[Xu, 2001] Xu, Y. (2001). Representation of reproducing kernels and the lebesgue constants on the ball. *Journal of Approximation Theory*, 112:295–310.

[Yu et al., 2015] Yu, Y., Wang, T., and Samworth, R. (2015). A useful variant of the Davis-Kahan theorem for statisticians. *Biometrika*, 102(2):315–323.
Appendix A: Useful results

Here we gather some of the results used throughout the paper.

**Lemma 10.** If $X$ is a $\mathbb{B}^d$-valued random variable distributed according to $F_\nu$, then $\|X\|^2$ follows a distribution $\text{Beta}(\frac{d}{2}, \nu + \frac{1}{2})$.

**Lemma 11** (Threshold graphon degree density). Let $W$ be the threshold graphon and $f$ the probability density function of $d_W(X)$, where $X \sim F_\nu$ we have for $t > 0$

$$f(t) = \frac{\tau^d}{(1 - I^{-1}(2t))^{\frac{d}{2}+1}} \left(1 - \frac{\tau^2}{1 - I^{-1}}\right)^{\nu - \frac{1}{2}} \frac{1}{I(2t)^{\frac{d}{2}+\nu}(1 - I(2t))^{-\frac{1}{2}}}$$

where we use the notation $I(t) = I_t(d + \nu, \frac{1}{2})$.

**Proof.** It is well known that the function $t \rightarrow I_t(a, b)$ is differentiable and it is straightforward to check that $g(t)$ is also differentiable for $t > 0$. Taking the derivative of $F_{d_W}(t) = I_{g(t)}(d + \nu, \frac{1}{2})$ the result follows from simple computations

$$f_g(t) = \frac{1}{B(d + \nu, \frac{1}{2})} \frac{\tau^d}{(1 - I^{-1}(2t))^{\frac{d}{2}+1}} \left(1 - \frac{\tau^2}{1 - I^{-1}}\right)^{\nu - \frac{1}{2}} \frac{1}{I(2t)^{\frac{d}{2}+\nu}(1 - I(2t))^{-\frac{1}{2}}}$$

**Lemma 12.** For $\tau \geq 0$ we have for $0 \leq t \leq 1$

$$d_W(tN) = \frac{1}{2} I_{\frac{\tau}{1 - \frac{\tau}{\nu}}} \left(\nu + \frac{d}{2}, \frac{1}{2}\right)$$

where $I_x(a, b) = \frac{1}{\text{Beta}(a, b)} \int_0^x t^{a-1}(1-t)^{b-1}$ is the regularized incomplete Beta function.

**Lemma 13.** Let $W(\langle x, y \rangle) = \mathbb{1}_{\{\langle x, y \rangle \geq \tau\}}$ be a graphon on $\mathbb{B}^d$, with $0 < \tau < 1$, and $F_\nu \in \mathcal{F}$. Then the function

$$h \rightarrow \mathcal{F}_\nu(\{x \in \mathbb{B}^d : d_W(x) \geq h\})$$

is continuous in $(0, h^*)$, where $h^* := \mathcal{F}_\nu(\mathbb{B}^d \setminus \tau \mathbb{B}^d)$.

**Proof.** By Lemma 12 we have

$$d_W(tN) = \frac{1}{2} I_{\frac{\tau}{1 - \frac{\tau}{\nu}}} \left(\nu + \frac{d}{2}, \frac{1}{2}\right)$$

from which we see that $t \rightarrow d_W(tN)$ is strictly increasing on $(\tau, 1]$ and its range is $(0, h^*)$. Then for any $h \in (0, h^*)$ there exists $t_h$ such that $\{x \in \mathbb{B}^d : d_W(x) \geq h\} = \mathbb{B}^d \setminus t_h \mathbb{B}^d$. Moreover, $t_h = \sqrt{\frac{\tau}{1 - I^{-1}(2h)}}$. It is easy to see that $h \rightarrow t_h$ is continuous on $(0, h^*)$ and given that $F_\nu$ is absolutely continuous with respect to the Lebesgue measure, we have that $h \rightarrow \mathcal{F}_\nu(\{x \in \mathbb{B}^d : d_W(x) \geq h\})$ is continuous in $(0, h^*)$.

The following result gives a characterization for a basis of $\mathcal{V}_t$. The proof can be found in [Dai and Xu, 2013, Thm.11.1.12]
Theorem 14. The space $V_n$ has a basis consisting on functions $G^w_{\gamma}(\langle x, \psi_i \rangle)$ for some points \{\psi_i\}_{1 \leq i \leq \dim(V)} \subset S^{d-1}.

Proposition 15. Let \{X^\mu_i\}_{1 \leq i \leq n} and \{X^\nu_i\}_{1 \leq i \leq n} be two sets of points distributed under $F_\mu$ and $F_\nu$ respectively for $\mu, \nu > 0$. Let $\tau$ be in $(0, 1]$ and assume that $\mu > \nu$, then we have
\[
P(\langle X^\mu_i, X^\mu_j \rangle \leq \tau) < P(\langle X^\nu_i, X^\nu_j \rangle \leq \tau)
\]
for $i \neq j$. Moreover, there exists $\tau' \in (0, 1]$ such that
\[
P(\langle X^\nu_i, X^\nu_j \rangle \leq \tau) = P(\langle X^\mu_i, X^\mu_j \rangle \leq \tau')
\]
for $i \neq j$.

Remark 5 (Case $\tau = 0$). It is easy to see that in the case $\tau = 0$ any measure with spherical symmetry we define the same $W$-random graph model. Intuitively speaking, the norm of the latent points is not used to decide the nodes connection, but only the fact that they belong to the same semisphere. In consequence, in the case $\tau = 0$ we cannot recover the measure (nor distributional information about the latent points) from the adjacency matrix alone.

Proposition 16. For the threshold graphon $W = W_\tau$ for $\tau \geq 0$ and $\{X_i\}_{1 \leq i \leq n} \sim F_\nu$ for $\nu > -1/2$, we have for any $1 \leq i \leq n$
\[
P(d_W(t_1N) \leq d_W(X_i) \leq d_W(t_2N)) = F_{Beta}(t_2^2) - F_{Beta}(t_1^2)
\]
where $0 < \tau < t_1 < t_2$ and $F_{Beta}(\cdot)$ is the cumulative distribution function of $Beta \left( \frac{d}{2}, \nu + \frac{1}{2} \right)$. In addition, we have that
\[
P(d_W(X_i) = 0) = F_{Beta}(\tau)
\]

A.1. Eigenvalue concentration

The following theorem is a slight reformulation of the [Bandeira and Van Handel, 2016, Cor.3.12]

Theorem 17 (Bandeira-Van Handel). Let $Y$ be a $n \times n$ symmetric random matrix whose entries $Y_{ij}$ are independent centered random variables. There exists a universal constant $C_0$ such that for $\alpha \in (0, 1)$
\[
P\left(\|Y\|_{op} \geq 3\sqrt{2D_0} + C_0\sqrt{\log n/\alpha}\right) \leq \alpha
\]
where $D_0 = \max_{0 \leq i \leq n} \sum_{j=1}^n Y_{ij}(1 - Y_{ij})$.

Using the previous theorem with $Y = \hat{T}_n - T_n$, which is centered and symmetric, we obtain the tail bound
\[
P\left(\|\hat{T}_n - T_n\|_{op} \geq 3\sqrt{2D_0} + C_0\sqrt{\log n/\alpha}\right) \leq \alpha
\]

The next theorem is proven in [Araya, 2020] and gives a finite sample bound for the individual eigenvalues of $T_n$ with respect to the eigenvalues of the integral operator $T_W$. 
Theorem 18. [Araya, 2020, Thm.2] Let $(\Omega, \mu)$ be a probability space and $W : \Omega^2 \rightarrow \mathbb{R}$ be a $L^2(\Omega^2)$ kernel. Let $|\lambda_1| \geq |\lambda_2| \geq \cdots$ be the eigenvalues of integral operator $T_W$ and $\{\phi_i\}_{i=1}^{\infty}$ the a set of orthonormal eigenfunctions. Assume that $\|\phi_i\|_\infty = \mathcal{O}(i^s)$ and $|\lambda_i| = \mathcal{O}(i^{-\delta})$, where $\delta > 2s + 1$. Then we have with probability larger than $1 - \alpha$

$$|\lambda_i(T_n) - \lambda_i| \lesssim i^{-\delta+2s+1}n^{-\frac{1}{2}} \quad \text{for } 1 \leq i \leq n$$

The following proposition gives a high probability bound for $\delta_2(\lambda(T_n), \lambda(T_W))$

Proposition 19. Let $W$ be a graphon on $\mathbb{B}^d$ of the form $W(x, y) = f(\langle x, y \rangle)$ and $f \in S^p_\nu([-1,1])$ for $p > 2\nu - 1 + \frac{5d}{2}$, then we have with probability larger than $1 - \alpha$

$$\delta_2(\lambda(T_n), \lambda(T_W)) \lesssim n^{-\frac{1}{2}}$$

Proof. Define $d_l := \dim (\mathcal{Y}_n)$. We will assume without loss of generality that $\{\lambda_k^*\}_{k \geq 0}$ is order decreasingly. Indeed, if $\sum_{l \geq 1} |\lambda_l^*| d_l < \infty$ holds, then $\{\lambda_k^*\}_{k \geq k_0}$ for some $k_0 \in \mathbb{N}$ large enough, given that $d_l \propto l^{d-1}$ (with means that there exists $c, C > 0$ such that $cl^{d-1} \leq d_l \leq C l^{d-1}$ for $l$ large enough). Define $l : \mathbb{N} \rightarrow \mathbb{N}$ to be the such that $\lambda_i = \lambda_{l(i)}^*$. From the relation

$$\sum_{l=0}^{l(i)-1} d_l \leq l(i) \leq \sum_{l=0}^{l(i)} d_l$$

which implies that $l(i) = \mathcal{O}(i^\frac{1}{d})$.

Given that $f \in S^p_\nu([-1,1])$ the eigenvalues $\lambda_i^*$ satisfy $\sum_{l \geq 1} |\lambda_l^*|^2 d_l (1 + \nu_l^p) < \infty$, where $\nu_l = l(l + 2\nu + d - 1)$. This implies that $|\lambda_l^*| = \mathcal{O}(l^{-\delta^*})$ with $\delta^* = p + \frac{d}{2} + \varepsilon$ and $\varepsilon > 0$. In consequence, we have $|\lambda_l| = \mathcal{O}(i^{-\delta})$, with $\delta := \frac{p + \varepsilon}{d} + \frac{1}{2}$. By Lemma 4, we have $\|p_{k,l}\|_\infty = \mathcal{O}(l^{2\nu_d - d - 1})$, which given that $d_l \propto l^{d-1}$, translate to $\|\phi_R\|_\infty = \mathcal{O}(R^{2\nu_d - 1 + \frac{1}{2}})$, for every $R \in \mathbb{N}$. Using Theorem 18, with $s = \frac{2\nu_d - 1}{2d} + \frac{1}{2}$ and $\delta = \frac{p + \varepsilon}{d} + \frac{1}{2}$, we obtain

$$|\lambda_i(T_n) - \lambda_i| \lesssim i^{-\delta+2s+1}n^{-1/2}, \quad \text{for } 1 \leq i \leq n$$

with probability larger than $1 - \alpha$. If $p > 2\nu - 1 + \frac{5d}{2}$ we the RHS of the previous inequality is summable, with respect to $i$, and the result follows.

For a graphon $W = f(\langle x, y \rangle)$ on $\mathbb{B}^d$, it is often useful to consider the sequence of eigenvalues of $T_W$ indexed with repetition. We will define as $\{\lambda_i^*(f)\}_{i \geq 0}$ the sequence of eigenvalues with repetitions, also ordered in the decreasing order for the absolute value. It is easy to see that each $\lambda_i^*(f)$ will appear $\dim(\mathcal{V}_l)$ times in $\{\lambda_i^*(f)\}_{i \geq 0}$ (if there exists $k$ such that $\lambda_i^* = \lambda_k^*$, then it will appear $\dim(\mathcal{V}_l) + \dim(\mathcal{V}_k)$ times).

Lemma 20. If $W$ is graphon on $\mathbb{B}^d$ such that $W(x, y) = f(\langle x, y \rangle)$ and $f \in S^p_\nu([-1,1])$ for $p > 2\nu - 1 + \frac{5d}{2}$, with eigenvalues $\{\lambda_i^*\}_{i \geq 0}$ (without repetition) and eigenfunctions $\{\phi_i\}_{i \geq 0}$. Define the $n \times n$ matrix with entries $(T_n^*)_{ij} := \frac{1}{n} \sum_{l=0}^{n-1} \lambda_i^*(X_l) \phi_i(X_j)$. Then we have

$$\|T_n - T_n^*\|_{op} = \mathcal{O}_\alpha\left(\frac{1}{\sqrt{n}}\right)$$

with probability larger than $1 - \alpha$. 
Proof. We have that

$$(T_n - T'_n)_{ij} = \frac{1}{n} \sum_{l \geq n} \lambda_l' \phi_l(X_i) \phi_l(X_j)$$

and by [Araya, 2020, Thm.1] we have with probability larger than $1 - \alpha$

$$\|T_n - T'_n\|_{op} = |\lambda_{n}'| + O_n(\frac{1}{\sqrt{n}})$$

On the other hand, given that $f \in S^p_{\nu}([-1, 1])$ for $p > 2\nu - 1 + \frac{5d}{2}$, we have that $\lambda_n \lesssim n^{-\frac{2\nu - 1}{2}}$, hence the conclusion follows. \hfill \Box

A.2. Eigenvectors concentration

We will use the following version of the Davis-Kahan sin $\theta$ theorem, which is stated and proved in [Yu et al., 2015]

Theorem 21 (Davis-Kahan). Let $\Sigma$ and $\hat{\Sigma}$ be two symmetric $\mathbb{R}^{n \times n}$ matrices with eigenvalues $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$ and $\hat{\lambda}_1 \geq \hat{\lambda}_2 \geq \cdots \geq \hat{\lambda}_n$ respectively. For $1 \leq r \leq s \leq n$ fixed, we assume that $\min \{\lambda_{r-1} - \lambda_r, \lambda_s - \lambda_{s-1}\} > 0$ where $\lambda_0 := \infty$ and $\lambda_{n+1} = -\infty$. Let $d = s - r + 1$ and $\bar{V}$ and $\hat{V}$ two matrices in $\mathbb{R}^{n \times d}$ with columns $(v_r, v_{r+1}, \cdots, v_s)$ and $(\hat{v}_r, \hat{v}_{r+1}, \cdots, \hat{v}_s)$ respectively, such that $\Sigma v_j = \lambda_j v_j$ and $\Sigma \hat{v}_j = \hat{\lambda}_j \hat{v}_j$. Then there exists an orthogonal matrix $O$ in $\mathbb{R}^{d \times d}$ such that

$$\|\bar{V}O - V\|_F \leq \frac{2^{3/2} \min \{\sqrt{d} \|\Sigma - \hat{\Sigma}\|_{op}, \|\Sigma - \hat{\Sigma}\|_F\}}{\min \{\lambda_{r-1} - \lambda_r, \lambda_s - \lambda_{s+1}\}}$$

(10)

We recall that $\Phi_k = \frac{1}{\sqrt{n}} (\phi_k(X_1), \phi_k(X_2), \cdots, \phi_k(X_n))^T$. For $k, k' \in \mathbb{N}$ such that $k' > k$, we define $\Phi_{k:k'}$ as the matrix with columns $\Phi_k, \Phi_{k+1}, \cdots, \Phi_{k'}$. Define $V_1(k, k') = \|\sum_{l=k}^{k'} \phi_l^2\|_\infty$.

Proposition 22. We have with probability at least $1 - \alpha$

$$\|\Phi_{k:k'} \Phi_{k:k'}^T - \text{Id}_{|k-k'|}\|_{op} \lesssim \frac{V_1(k, k')}{n} \wedge \frac{\sqrt{V_1(k, k')}}{n}$$

Proof. The proof is identical to the proof of [Araya, 2020, Prop.4], which uses Matrix Bernstein inequality [Tropp, 2012, Thm.6.1]. \hfill \Box

Lemma 23. Let $B$ a $n \times d$ matrix with full column rank. Then we have

$$\|BB^T - B(B^T B)^{-1}B^T\|_F = \|\text{Id}_d - B^T B\|_F$$

Proof. We have

$$\|BB^T - B(B^T B)^{-1}B^T\|_F = \|B((B^T B)^{-1} - \text{Id}_d)B^T\|_F$$

and by definition of the Frobenious norm and cyclic property of the trace

$$\|B((B^T B)^{-1} - \text{Id}_d)B^T\|_F^2 = \text{tr}(B((B^T B)^{-1} - \text{Id}_d)B^T B((B^T B)^{-1} - \text{Id}_d)B^T)$$

$$= \text{tr}((\text{Id}_d - B^T B)^2)$$

$$= \|\text{Id}_d - B^T B\|_F^2$$

\hfill \Box
Appendix B: Proofs

Here we gather the proofs of the main results of the article.

**Proof of Lemma 10.** It is classic (see [Kelker, 1970, Sec.5]) that for a spherically symmetric distribution with density of the form $p(y) = g(\|y\|^2)$ where $y \in \mathbb{B}^d$, then the norm will have density $h(r) = \frac{2\pi^{d/2}}{\Gamma(d/2)} r^{d-1} g(r^2)$. The c.d.f for the radius of variable distributed following $F_{\nu}$ is proportional to $\int_0^t r^{d-1}(1 - r^2)^{\nu - \frac{1}{2}} dr$ using the change of variables $r \to r^2$ we obtain that the square of the norm have density $\int_0^t r^{d-1}(1 - r^2)^{\nu - \frac{1}{2}} dr$ where we recognize the density of a Beta$(\frac{d}{2}, \nu + \frac{1}{2})$.

**Proof of Prop. 16.** Notice that in the case of the threshold graphon, the degree function $t \to d_W(tN)$ is increasing. Using this we have that

\[
P(d_W(t_1N)) \leq d_W(X_i) \leq d_W(t_2N) = P(\|X_i\| \in [t_1, t_2])
\]

Using the previous and Lemma 10, the result follows.

**Proof of Prop. 9.** We will assume that $h'$ is a rational number. We choose $k, n' \in \mathbb{N}$ such that $h = \frac{k}{n'}$. We put $\theta_1 = 2.5$. We saw in Sec. that $P(\frac{k-1}{n'} \leq \tilde{d}_f(x) \leq \frac{k}{n'}) \lesssim \frac{(k/n')^{\theta_1}}{\sum_{i=k+1}^{n'} i^{-\theta_1}}$. We have following claim. **Claim 1:** There exist a constant such that $P(\tilde{d}_f(x) \geq h') \lesssim Ch'^{1-\theta_1}$. We proof this claim. We have

\[
P(\tilde{d}_f(x) \geq h') \lesssim \sum_{i=k+1}^{n'} P(i - 1 \leq \tilde{d}_f(x) \leq i)
\]

\[
\lesssim \frac{1}{(n')^{-\theta_1}} \sum_{i=k+1}^{n'} i^{-\theta_1}
\]

\[
\lesssim \frac{1}{(n')^{-\theta_1}} \sum_{i=k}^{n'} i^{-\theta_1}
\]

\[
\lesssim \frac{k}{n'}^{1-\theta_1} = h'^{1-\theta_1}
\]

We have the following: **Claim 2:** There exists a linear function $L_\alpha$ such that $\tilde{d}_f(x) \geq L_\alpha(d_f(x))$, we prove this claim. Define

\[
J_\alpha = \int_{\mathbb{B}^d \setminus B^{d}(0, \sqrt{\alpha})} \frac{1}{y_1} dF_{\nu}(y)
\]

\[
\kappa_\alpha = \int_{\mathbb{B}^d(0, \sqrt{\alpha})} dF_{\nu}(y) / J_\alpha
\]

By definition, we have

\[
\frac{d_f(x)}{J_\alpha} = \frac{1}{J_\alpha} \int_{\mathbb{B}^d \setminus B^{d}(0, \sqrt{\alpha})} \frac{\alpha}{x_1^2 y_1^2} \wedge 1 dF_{\nu}(y) + \kappa_\alpha
\]
By definition, \( \tilde{d}_f(x) \) is larger than the first term in the RHS in the previous expression. This implies that
\[
\tilde{d}_f(x) \geq \frac{1}{J_\alpha} d_f(x) - \kappa_\alpha
\]
Defining \( L_\alpha(x) = x/J_\alpha - \kappa_\alpha \), the claim follows.

With this, we have that there exist a constant \( C > 0 \) such that
\[
P(d_f(x) \geq h) \leq P(\tilde{d}_f(x) \geq L_\alpha(h)) \]
\[
\leq C(L_\alpha(h))^{1-\theta_1} \]
\[
= C(\frac{h}{J_\alpha} - \kappa_\alpha)^{1-\theta_1} \]
\[
\leq C(\frac{1}{J_\alpha} - \kappa_\alpha)^{1-\theta_1} \]
which proves the proposition.

**Proof of Lemma 12.** For \( t \leq \tau \) we have that \( tN \in B^d(0, \tau) \), which implies that \( d_W(tN) = 0 \). The result for this case follows by noting that \( I_0(a,b) = 0 \) for any \( 0 \leq a, b \leq 1 \). For \( t > \tau \), call \( h = (\tau/t)^2 \) we have
\[
d_W(tN) = \int_{B^d} 1_{(tN,y) \geq \tau}(1 - \|y\|^2)^{\nu - \frac{1}{2}}
\]
\[
\propto \int_{h}^{1} \int_{0}^{1} \sqrt{1 - x_1^2} r^{d-2}(1 - x_1^2 - r^2)^{\nu - \frac{1}{2}} dr dx_1
\]
\[
\propto \int_{h}^{1} (1 - x_1)^{\frac{d}{2} + \nu - 1} dx_1 \int_{0}^{1} (1 - t)^{\nu - \frac{1}{2}} t^{\frac{d-3}{2}} dt
\]
\[
\propto \int_{0}^{1-h} x_1^{\frac{d}{2} + \nu - 1} \frac{1}{x_1^2} dx_1
\]
where we did a change of variables \( t = \frac{r^2}{1 - x_1^2} \) in the third line. The result follows from the fact the both quantities integrate 1.

**Proof of Lemma 4.** From [Dai and Xu, 2013][Thm.11.1.12] we know that for each \( n \) there exists \( \psi_k \in S^{d-1} \) such that \( G_n^{\gamma_n}(\psi_k, x) \) is a basis of \( \mathcal{Y}_n \). We take \( p_{k,n}(x) = G_n^{\gamma_n}(\psi_k, x) \) for \( 1 \leq k \leq \dim(\mathcal{Y}_n) \). From [Dai and Xu, 2013, Eq.B.2.2] and [Szego, 1939, Thm.7.32.1] we have \( \|p_{k,n}\|_{\infty} \leq |G_n^{\gamma_n}(1)| \times n^{\gamma_n} \), because Gegenbauer polynomials are Jacobi polynomials with the repeated exponent parameter. For the second inequality, we use (4) and (2) to obtain
To prove the second assertion, we see that by routine computations, the density of the inner product conditional to the latent points, to obtain with probability larger than $\alpha/(n-1)$.

$\sum_{k=1}^{\dim(Y_n)} p_{k,n}(x) = c_n \frac{n+\gamma n}{\gamma} \int_{-1}^{1} G_{n,\nu}^\nu (\langle x, y \rangle + \sqrt{1-\|x\|^2} \sqrt{1-\|y\|^2} t) (1-t^2)^{\nu-1} dt$

$\leq \|G_{n,\nu}^\nu\|_{\infty} c_n \frac{n+\gamma n}{\gamma} \int_{-1}^{1} (1-t^2)^{\nu-1} dt$

$\lesssim |G_{n,\nu}^\nu(1)| n c_n \int_{-1}^{1} (1-t^2)^{\nu-1} dt$

$\lesssim n^{2\nu-d-1} c_n \int_{-1}^{1} (1-t^2)^{\nu-1} dt$

Proof of Prop. 15. For every $i$, we can write $X_i^\mu = R_i^\mu U_i$, where $R_i$ and $U_i$ are independent and $U_i$ is uniformly distributed on $S^{d-1}$. Similarly, we can decompose $X_i^\nu = R_i^\nu V_i$, where $V_i$ is is uniformly distributed on $S^{d-1}$. Given that $\mu < \nu$ we have that $(1-\|x\|^2)^{\nu-\frac{1}{2}} < (1-\|x\|^2)^{\mu-\frac{1}{2}}$, for $x \in B^d$. This implies that $P(R_i^\nu < \tau) > P(R_i^\mu < \tau)$. Given that $\langle X_i^\mu, X_i^\nu \rangle = R_i^\mu R_i^\nu \langle U_i, U_j \rangle$ and $\langle X_i^\nu, X_j^\nu \rangle = R_i^\nu R_j^\nu \langle V_i, V_j \rangle$, and $\langle U_i, U_j \rangle \overset{D}{=} \langle V_i, V_j \rangle$ (they are equal in distribution), it is easy to see that

$P(R_i^\mu R_j^\nu \langle U_i, U_j \rangle \leq \tau) < P(R_i^\nu R_j^\nu \langle V_i, V_j \rangle \leq \tau)$

To prove the second assertion, we see that by routine computations, the density of the inner product $\langle X_i^\mu, X_i^\mu \rangle$ is

$t \to \int_{0}^{1} \int_{0}^{1} ((sr)^2 - t^2)^{\frac{4}{2} - 1}(1-t^2)^{\mu-1}(1-s^2)^{\mu-1} 1_{sr > |t|} dr ds$

Proof of Prop. 2. Conditional to $X_i$, the random variable $d_G(X_i)$ is a sum of independent random variables, hence by the strong law of large number $d_G(X_i) \to d_W(X_i)$ almost surely. By the continuity of the function $t \to \frac{\tau}{\sqrt{1-I^{-1}(t)}}$, we deduce that

$\frac{\tau}{\sqrt{1-I^{-1}(2d_G(X_i)/(n-1))}} \to \frac{\tau}{\sqrt{1-I^{-1}(2d_W(X_i))}} = \|X_i\|$

in the almost sure sense.

Proof of Lemma 6. Invoke Theorem 17 with $Y = \hat{T}_n - T_n$, which has independent centered entries conditional to the latent points, to obtain with probability larger than $1 - \alpha/2$

$\|\hat{T}_n - T_n\|_{op} \lesssim \alpha \frac{1}{\sqrt{n}}$

because $D_0 = \max_{0 \leq i < n} \sum_{j=1}^{n} \Theta_{ij}(1-\Theta_{ij})$ is $O(n \rho_n)$, by the definition of $\Theta$. Thus, there exists $n_0' \in N$ such that for all $n \geq n_0$ we have $\|\hat{T}_n - T_n\|_{op} \leq \frac{A^2}{2^p \sqrt{d}}$. It is easy to see that in this case $n_0' = O(\Delta^p \sqrt{d} \log 2/\alpha)$. 

From Theorem 19 we have that, there exists $n''_0$ such that for $n \geq n''_0$ we have

$$\delta_2\left(\lambda(T_n), \lambda(T_W)\right) \lesssim_\alpha \frac{1}{\sqrt{n}} \leq \frac{\Delta^*}{4},$$

with probability larger than $1 - \alpha/2$. We see here that $n''_0 = \mathcal{O}(\Delta^{*-1} \log 1/\alpha)$. Taking $n_0 = \max\{n'_0, n''_0\}$ we have that $\mathbb{P}(\mathcal{E}) \geq \alpha/2$, for $n \geq n_0$. \hfill \Box

**Proof of Prop. 7.** First, notice that under $\mathcal{E}(W, n)$, we have

$$\|T_n - \hat{T}_n\|_{op} \leq \frac{\Delta^*}{4\sqrt{d}} \leq \frac{\Delta^*}{4}$$

because $\Delta^* \leq 1$, given that $0 \leq W \leq 1$. We also have $\delta_2(\lambda(T_n), \lambda(T_{W})) \leq \frac{\Delta^*}{4}$. From that and the definition of $\delta_2(\cdot, \cdot)$ we deduce that there are at least $d$ eigenvalues of $T_n$ at distance less than $\Delta^*/4$ from $\lambda_1^*$ (given the multiplicity of $\lambda_1^*$). But each eigenvalue of $T_n$ is at distance at most $\Delta^*/4$ from an eigenvalue of $T_W$, and given that $dist(\lambda_1^*, \lambda(T_W) \setminus \{\lambda_1^*\}) = \Delta^*$ we have that there are exactly $d$ eigenvalues of $T_n$ at distance at most $\Delta^*/4$ from $\lambda_1^*$. By the triangle inequality and $\|T_n - \hat{T}_n\|_{op} \leq \frac{\Delta^*}{4}$ we deduce that there exists a set of exactly $d$ eigenvalues of $\hat{T}_n$ at distance at most $\frac{\Delta^*}{4}$ from $\lambda_1^*$. \hfill \Box

**Proof of Thm. 8.** By Prop. 7 we know that, under the eigengap condition, there is a cluster $\hat{\Lambda}_1$ of exactly $d$ eigenvalues of $\hat{T}_n$ and, another cluster $\Lambda_1$ of $d$ eigenvalues of $T_n$, such that all the elements in both clusters are at distance at most $\Delta^*/2$ from $\lambda_1^*$. We called $V$ (resp. $\hat{V}$) to the $n \times d$ matrix, where the columns are the eigenvectors of $T_n$ (resp. $\hat{T}_n$) associated with $\Lambda_1$ (resp. $\hat{\Lambda}_1$). By Theorem 17 we have that

$$\|\hat{T}_n - T_n\|_{op} \lesssim_\alpha \frac{1}{\sqrt{n}}$$

with probability larger than $1 - \alpha$. By Thm. 21 we have that with probability larger than $1 - \alpha$

$$\|VV^T - \hat{V}\hat{V}^T\|_{op} \lesssim_\alpha \frac{\sqrt{d}}{\Delta^* \sqrt{n}}$$

We will assume that $\{\lambda_i\}_{i \geq 0}$ is the sequence of eigenvalues of $T_{W}$ indexed with repetition. To prove the theorem will be sufficient to show that $\|\Phi_t d^d \Phi_t^T - VV^T\|_{op} = \mathcal{O}(\frac{1}{\sqrt{n}})$ with probability at least $1 - \alpha$.

We define the matrices $T'_n = \frac{1}{n} \sum_{i=0}^{n-1} \lambda_i \Phi_i \Phi_i^T$ and $\hat{T}_n = \frac{1}{n} \sum_{i=0}^{n-1} \lambda_i \hat{\Phi}_i \hat{\Phi}_i^T$, where the vectors $\{\Phi_0, \cdots, \Phi_{n-1}\}$ are obtained from $\Phi_0, \cdots, \Phi_{n-1}$ by a Gram-Schmidt orthonormalization process. Observe that $\hat{T}_n \hat{\Phi}_i = \lambda_i \hat{\Phi}_i$. We have the following claim.

**Claim 1:** with probability larger than $1 - \alpha$ we have $\|\hat{T}_n - T'_n\|_{op} \lesssim_\alpha \mathcal{O}(\frac{1}{\sqrt{n}})$.

Assume this claim for the moment and define $\hat{V}$ a matrix with columns $\hat{\Phi}_1, \cdots, \hat{\Phi}_d$. By Lemma 20 we have $\|T_n - T'_n\|_{op} = \mathcal{O}(\frac{1}{\sqrt{n}})$ with probability larger than $1 - \alpha$, which implies that $\|\hat{T}_n - T'_n\|_{op} \lesssim_\alpha \frac{1}{\sqrt{n}}$ (by triangle inequality and Claim 1) and by Thm. 21 we have that

$$\|\hat{V}\hat{V}^T - VV^T\|_F \lesssim_\alpha \frac{\sqrt{d}}{\Delta^* \sqrt{n}}$$
We will now prove Claim 1. Consider the notation

\[ \tilde{V}_{d_l} := (\tilde{\Phi}_{d_l} | \tilde{\Phi}_{d_{l+1}} | \cdots | \tilde{\Phi}_{d_{l+1}}) \]

\[ \Phi_{d_l} := \Phi_{d_{l:d_{l+1}}} = (\Phi_{d_l} | \Phi_{d_{l+1}} | \cdots | \Phi_{d_{l+1}}) \]

Given that \( \tilde{\Phi} \) is obtained by a Gram-Schmidt process from \( \Phi \), we have that \( \text{span}(\tilde{V}_{d_l}) = \text{span}(\Phi_{d_l}) \), where \( \text{span}(A) \) is the linear span of the columns of matrix \( A \). Hence the orthogonal projectors \( \tilde{V}_{d_l} \tilde{V}_{d_l}^T \) and \( \Phi_{d_l} (\Phi_{d_l}^T \Phi_{d_l})^{-1} \Phi_{d_l}^T \) are equal for every \( l \leq l(n) \), where \( l(n) \) is defined by \( l(n) = \min\{l' \in \mathbb{N} : \sum_{i=0}^{l'} d_i \leq n \} \).

On the other hand, we have that with probability at least \( 1 - 2\alpha \)

\[
\| \Phi_{d_l} (\Phi_{d_l}^T \Phi_{d_l})^{-1} \Phi_{d_l}^T - \Phi_{d_l} \|_F = \| \Phi_{d_l} \Phi_{d_l} - \text{Id}_{d_l} \|_F 
\leq \alpha \sqrt{\frac{\nu_1(d_l, d_{l+1})}{n}}
\]  

where we used Lemma 23 in the first step and Prop. 22, together with the bound \( \|A\|_F \leq \sqrt{d_l} \|A\|_{op} \) for a matrix of size \( d_l \), in the last step. Notice that is possible to use Lemma 23 because with probability \( 1 - \alpha \) we have \( \| \Phi_{d_l}^T \Phi_{d_l} - \text{Id}_{d_l} \|_{op} \leq \alpha \frac{\nu_1(d_l, d_{l+1})}{n} \), and \( \frac{\nu_1(d_l, d_{l+1})}{n} \) is defined by \( l(n) = \min\{l' \in \mathbb{N} : \sum_{i=0}^{l'} d_i \leq n \} \). Hence, the event that \( \Phi_{d_l} \) has full rank has probability at least \( 1 - \alpha \). By Lemma 4 we have that \( \nu_1(d_l, d_{l+1}) = O(l^{2\nu - 1 - d}) \), but given the assumption on the Sobolev regularity of \( f \), we have \( \sum_{l=0}^{l(n)} |\lambda_l^*| \sqrt{d_l \nu_1(d_l, d_{l+1})} = O(1) \), for all \( l \leq l(n) \). Indeed, we have that \( \sqrt{d_l \nu_1(d_l, d_{l+1})} = O(l^{2\nu - 1 + d}) \) and \( |\lambda_l^*| = O(l^{\delta^*}) \), where \( \delta^* > (2\nu - 1 + 3d) \), which implies that \( |\lambda_l^*| \sqrt{d_l \nu_1(d_l, d_{l+1})} = O(l^{-2d}) \), which is summable. Given the spectral expansion of \( \tilde{T}_n \) and \( T'_n \), we have

\[
\| \tilde{T}_n - T'_n \|_{op} \leq \sum_{l=0}^{l(n)} |\lambda_l^*| \| \Phi_{d_l} (\Phi_{d_l}^T \Phi_{d_l})^{-1} \Phi_{d_l}^T - \Phi_{d_l} \Phi_{d_l}^T \|_{op}
\]

Bounding the operator norm by the Frobenius norm and (12), we have

\[
\| \tilde{T}_n - T'_n \|_{op} \leq \sum_{l=0}^{l(n)} \lambda_l^* \sqrt{\frac{\nu_1(d_l, d_{l+1})}{n}} \leq \alpha \frac{1}{\sqrt{n}}
\]

This proves Claim 1. Notice that by (12) we have that \( \| \tilde{V} \Phi_{d_l}^T - \tilde{V} \tilde{V}^T \|_{op} \leq \alpha, \) which by triangular inequality gives that

\[
\| \tilde{V} \tilde{V}^T - \tilde{V} \Phi_{d_l}^T \|_F \leq \alpha, \frac{1}{\Delta^* \sqrt{n}}
\]

which concludes the proof. \( \square \)