CAN SMOOTH GRAPHONS IN SEVERAL DIMENSIONS BE REPRESENTED BY SMOOTH GRAPHONS ON $[0,1]$?

SVANTE JANSON AND SOFIA OLHEDE

ABSTRACT. A graphon that is defined on $[0,1]^d$ and is Hölder(α) continuous for some $d \geqslant 2$ and $\alpha \in (0,1]$ can be represented by a graphon on $[0,1]$ that is Hölder(α/d) continuous. We give examples that show that this reduction in smoothness to α/d is the best possible, for any d and α; for $\alpha = 1$, the example is a dot product graphon and shows that the reduction is the best possible even for graphons that are polynomials.

A motivation for studying the smoothness of graphon functions is that this represents a key assumption in non-parametric statistical network analysis. Our examples show that making a smoothness assumption in a particular dimension is not equivalent to making it in any other latent dimension.

1. Introduction

Networks or graphs are a convenient and parsimonious data structure for representing objects and their interactions. Initial interest in networks in statistics has focussed on fitting simple and parametric models to summarize data structure [10], such as the Chung-Lu or expected degree model, or variants of the stochastic block model. What most statistical network models satisfy is a probabilistic invariance to permutations, and this invariance leads to a natural representation of a graph generating mechanism via a graphon or a graph limit function [11] via the Aldous–Hoover theorem.

In general, a graphon can be defined on any probability space $\mathcal{S} = (\mathcal{S}, \mathcal{F}, \mu)$. A *graphon* on $\mathcal{S}$ is a symmetric measurable function $W : \mathcal{S}^2 \to [0,1]$. As is well known, graphons representing a graph limit or a random graph model are not unique, and there is the notion of (weak) equivalence of graphons; see [11]. In particular, any graphon is equivalent to a graphon defined on $[0,1]$; thus from an abstract point of view, it suffices to consider this case, and indeed, several papers consider only such graphons. However, in applications, it is often useful to consider other spaces $\mathcal{S}$, since not all models of networks are naturally formulated in terms of a graphon on $[0,1]$. In particular, it is often natural to use subsets of $\mathbb{R}^d$ with $d \geqslant 2$; some examples are the random dot product model [2], and applications where the latent dimension is interpreted as a position in a social space, cf. [8]. We consider below the case $\mathcal{S} = [0,1]^d$ (with Lebesgue measure); this means that each node is assigned d latent variables, which are independent and uniformly distributed on $[0,1]$.

From a statistical perspective, thus at best we can only estimate an element of the equivalence class of a graphon, just like in statistical shape analysis, where we

Date: 19 January, 2021; revised 1 February 2021.

1991 *Mathematics Subject Classification.* 05C80; 62G05.

Supported by the Knut and Alice Wallenberg Foundation.

Supported by the European Research Council under Grant CoG2015-682172NETS.

may estimate a shape but we have to factor out shifts and rotations, as they do not alter the underlying shape [4]. The probabilistic invariance that is most closely studied in statistics is an invariance to temporal and spatial shifts, most commonly found in stochastic processes [1]. The graphon function by analogy can therefore be compared to the spectral density of a stochastic process, except it does not permit as easy a characterisation as the spectral density of a random field or time series. Despite this fact, to enable estimation in statistics assumptions of regularity of a graphon, such as Lipschitz or Hölder continuity, has become common when analysing networks non-parametrically [7; 13; 14].

Despite the recent progress in statistics, machine learning and network data analysis, it is unclear how restrictive the assumption of either a Lipschitz or Hölder(α) graphon on $[0,1]$ is, and also how it compares with such assumptions for graphons defined on other probability spaces $\mathcal{S}$.

The aim of this paper is to explore the consequences of assuming smoothness of a graphon when its arguments takes values in $[0,1]^d$. It is easy to see that any Hölder(α) graphon on $[0,1]^d$ ($d \geqslant 2$) is equivalent to a Hölder(α/d) graphon on $[0,1]$; in particular any Lipschitz smooth graphon on $[0,1]^d$ is equivalent to a Hölder($1/d$) graphon on $[0,1]$ (Theorem 2.1). Moreover, we give examples showing that in general this is the best possible. In particular, we exhibit a simple infinitely differentiable graphon on $[0,1]^d$ that is not equivalent to any Hölder(q) graphon on $[0,1]$ for $q > 1/d$. The interpretation of this is that a smoothness assumption in a particular dimension "has teeth" and thus represents a real restriction, which furthermore depends on the dimension.

What is the statistical importance of that result? By assuming smoothness we are able to exhibit a member of the equivalence class of graphon functions and so bound any approximation error going from a block model to a Hölder(α) smooth function, drawing on classical results in numerical analysis [6] and the convergence of order statistics, see e.g. [13]. Furthermore by averaging we reduce variance and so make the average block heights nicely behaved random variables (controlled tail behaviour), irrespectively of what groupings we keep in a block. So the urge to average is natural, as so many nice results come from this act. However it comes at a price, namely to justify averaging in blocks we need to assume graphon smoothness in $[0,1]$ and not all graphons will satisfy this assumption.

2. Notation and main results

Recall that, for a given $\alpha \in (0,1]$, a function f defined on a subset $\mathcal{S}$ of a Euclidean space $\mathbb{R}^d$, say, is *Hölder*(α) if there exists a constant $C < \infty$ such that

$$|f(x) - f(y)| \leqslant C|x-y|^\alpha, \qquad x, y \in \mathcal{S}. \tag{2.1}$$

Functions that are Hölder(1) are also called *Lipschitz*. In particular, this notion applies to graphons W defined on $[0,1]^d$; recall that then W is a function on $[0,1]^{2d}$.

As said above, graphons are not unique, see e.g. [11] and [9]. In particular, if $\mathcal{S}_1$ and $\mathcal{S}_2$ are two probability spaces and $\varphi : \mathcal{S}_1 \to \mathcal{S}_2$ is a measure-preserving map, then for any graphon W on $\mathcal{S}_2$, its *pull-back* $W^\varphi(x,y) := W\bigl(\varphi(x), \varphi(y)\bigr)$ is a graphon on $\mathcal{S}_1$ that is equivalent to W. The converse does not hold, but it holds "almost", see Proposition 3.1 below.

We note first a simple result showing that every Hölder continuous graphon on $[0,1]^d$ is equivalent to a graphon on $[0,1]$ that is Hölder continuous albeit with a different Hölder exponent after the change of dimension. We regard $[0,1]$ and $[0,1]^d$ as probability spaces equipped with the usual Lebesgue measure.

Theorem 2.1. *Let W be a graphon on $[0,1]^d$ that is Hölder(α) for some $d \geqslant 2$ and $\alpha \in (0,1]$. Then there exists an equivalent graphon on $[0,1]$ that is Hölder(α/d).*

In particular, if W is differentiable, or just Lipschitz, then there exists an equivalent graphon on $[0,1]$ that is Hölder$(1/d)$.

Proof. Several standard constructions of Peano curves yield a measure-preserving map $\varphi : [0,1] \to [0,1]^d$ that is Hölder$(1/d)$, see e.g. [12; 15; 5]. Then the pull-back W^φ is a graphon on $[0,1]$ that is equivalent to W and is Hölder$(1/d)$. □

Our main purpose is to show that Theorem 2.1 is the best possible, by exhibiting graphons, for which the exponent α/d cannot be improved.

Example 2.2. Let W be the graphon associated with the random dot product graph on $[0,1]^d$ given by

$$W(x,y) = ax \cdot y = c(x_1y_1 + \dots + x_dy_d), \qquad x,y \in [0,1]^d, \tag{2.2}$$

where $a > 0$ is a constant, $\cdot$ is the scalar product, and $x = (x_1,\dots,x_d)$, $y = (y_1,\dots,y_d)$. (The constant a may be chosen as $1/d$ to make $0 \leqslant W \leqslant 1$.)

W is a polynomial and thus infinitely differentiable. We show in Section 3 that W is not equivalent to any graphon on $[0,1]$ that is Hölder(α) for any $\alpha > 1/d$; in particular not to any Lipschitz or differentiable graphon on $[0,1]$. □

Example 2.3. Let $d \geqslant 2$ and $\alpha \in (0,1)$. Let $h_\alpha(t)$ be the Weierstrass function given by the lacunary Fourier series

$$h_\alpha(t) := \sum_{k=0}^{\infty} 2^{-k\alpha} \cos\bigl(2\pi 2^k t\bigr). \tag{2.3}$$

Then h_α is a symmetric and periodic real-valued function on $\mathbb{R}$. Furthermore, it is easy to see that $h_\alpha \in$ Hölder(α), see [16, Theorem II.(4.9)].

Now define, for $x = (x_1,\dots,x_d)$ and $y = (y_1,\dots,y_d)$ in $[0,1]^d$,

$$W(x,y) := \frac{1}{2} + a \sum_{i=1}^{d} h_\alpha(x_i - y_i), \tag{2.4}$$

where $a > 0$ is chosen so small that $0 \leqslant W(x,y) \leqslant 1$. Then W is a graphon on $[0,1]^d$, and W is Hölder(α) since h_α is. By Theorem 2.1, there exists a graphon W' on $[0,1]$ that is equivalent to W and which is Hölder(α/d). We show in Section 3 that this is the best possible; W is not equivalent to any graphon on $[0,1]$ that is Hölder(β) for any $\beta > \alpha/d$. □

3. Proofs

We first quote the following characterization of equivalence of graphons, proved by Borgs, Chayes and Lovász [3]. (See also [11, Theorem 13.10] and [9, Theorems 8.3 and 8.4].)

Proposition 3.1 (Borgs, Chayes and Lovász [3])**.** *Two graphons W_1 and W_2, defined on probability spaces $\mathcal{S}_1$ and $\mathcal{S}_2$, respectively, are equivalent if and only if there exists a third graphon W on a probability space $\mathcal{S}$ and two measure-preserving maps $\varphi_1 : \mathcal{S}_1 \to \mathcal{S}$ and $\varphi_2 : \mathcal{S}_2 \to \mathcal{S}$ such that W_j a.e. equals the pull-back W^{φ_j}, $j = 1,2$.*

Our proofs are based on a functional of graphons, defined as follows. Let $q > 0$. For a graphon W on a probability space $\mathcal{S}$, or more generally any measurable function $W : \mathcal{S}^2 \to \mathbb{R}$, we define

$$\Psi_q(W) := \int_{\mathcal{S}}\int_{\mathcal{S}}\Bigl(\int_{\mathcal{S}} |W(x,z)-W(y,z)|\,\mathrm{d}\mu(z)\Bigr)^{-q}\,\mathrm{d}\mu(x)\,\mathrm{d}\mu(y) \leqslant \infty. \tag{3.1}$$

This functional is related to integrals used for, e.g., L^p-versions of Hölder continuity, but note the negative power; thus $\Psi_q(W)$ is large (or infinite) when W is sufficiently smooth, and $1/\Psi_q(W)$ may be regarded as a special kind of measure of (lack of) smoothness.

The claims in the examples will follow from the lemmas below.

Lemma 3.2. *If W and W' are two equivalent graphons, possibly defined on different probability spaces, then $\Psi_q(W) = \Psi_q(W')$ for every $q > 0$.*

Proof. By Proposition 3.1, it suffices to prove this when W' is a.e. equal to a pull-back of W by a measure-preserving map. This case follows by trivial changes of variables in the integrals. ◻

Lemma 3.3. *If $\alpha > 0$ and W is a graphon on $[0,1]$ such that W is Hölder(α), then $\Psi_q(W) = \infty$ for every $q \geqslant 1/\alpha$.*

Proof. By assumption, $|W(x,z)-W(y,z)| \leqslant C|x-y|^\alpha$, and thus $\int_0^1 |W(x,z)-W(y,z)|\,\mathrm{d}z \leqslant C|x-y|^\alpha$. Hence, (3.1) yields

$$\Psi_q(W) \geqslant C^{-q}\int_0^1\int_0^1 |x-y|^{-q\alpha}\,\mathrm{d}x\,\mathrm{d}y = \infty, \tag{3.2}$$

since $q\alpha \geqslant 1$. ◻

Lemma 3.4. *If $d \geqslant 1$, $\mathcal{S} = [0,1]^d$ and $W(x,y) := ax \cdot y$ for $x, y \in [0,1]^d$ and some $a > 0$, then $\Psi_q(W) < \infty$ for every $q < d$.*

Proof. By homogeneity, we may without loss of generality assume $a = 1$. Then

$$|W(x,z)-W(y,z)| = |x\cdot z - y\cdot z| = |(x-y)\cdot z|. \tag{3.3}$$

Define, for $x \in \mathbb{R}^d$,

$$h(x) := \int_{[0,1]^d} |x\cdot z|\,\mathrm{d}z. \tag{3.4}$$

Then $h(x)$ is a continuous function of x, and $h(x) > 0$ for $x \neq 0$. Hence, $c_d := \inf\{h(x) : |x| = 1\} > 0$ by compactness of the unit sphere. Furthermore, homogeneity yields $h(x) \geqslant c_d|x|$ for every $x \in \mathbb{R}^d$. Consequently, using (3.3),

$$\int_{[0,1]^d} |W(x,z)-W(y,z)|\,\mathrm{d}z = h(|x-y|) \geqslant c_d|x-y|, \tag{3.5}$$

and the definition (3.1) yields

$$\Psi_q(W) \leqslant c_d^{-q}\int_{[0,1]^d}\int_{[0,1]^d} |x-y|^{-q}\,\mathrm{d}x\,\mathrm{d}y < \infty, \tag{3.6}$$

recalling the assumption $q < d$. ◻

Proof of claim in Example 2.2. Suppose that W' is a graphon on $[0,1]$ that is equivalent to W and also is Hölder(α) for some $\alpha > 1/d$. Take $q := 1/\alpha < d$. Then $\Psi_q(W') = \infty$ by Lemma 3.3 and $\Psi_q(W) < \infty$ by Lemma 3.4, which contradicts Lemma 3.2. □

Lemma 3.5. *If $d \geqslant 2$, $0 < \alpha < 1$ and W is given by* (2.4)*, then $\Psi_q(W) < \infty$ for every $q < d/\alpha$.*

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

$$|x-y|_\circ := \min\bigl(|x-y|, 1-|x-y|\bigr) \tag{3.7}$$

and define, more generally, for $x = (x_i)_1^d \in [0,1]^d$ and $y = (y_i)_1^d \in [0,1]^d$,

$$|x-y|_\circ := \sum_{i=1}^d |x_i - y_i|_\circ. \tag{3.8}$$

(These can be regarded as metrics on $\mathbb{T}$ and $\mathbb{T}^d$, where the unit circle $\mathbb{T}$ is regarded as $[0,1]$ with the endpoints 0 and 1 identified.)

The function h_α satisfies for some $c_1, c_2 > 0$ and all $y \in (0,1)$,

$$\|h(\cdot) - h(\cdot - y)\|_{L^1[0,1]} \geqslant c_1 \|h(\cdot) - h(\cdot - y)\|_{L^2[0,1]} \geqslant c_2 |y|_\circ^\alpha, \tag{3.9}$$

where the first inequality is a general property of lacunary series [16, Theorem V.(8.20)] and the second follows by Parseval's relation and a simple calculation which we omit. Consequently, (2.4) yields, using (3.9) in the last line, for any $y = (y_i)_1^d \in [0,1]^d$,

$$\begin{aligned}
\int_{[0,1]^d} \bigl|W(x,z) - W(y,z)\bigr| \,\mathrm{d}z &= a \int_{[0,1]^d} \left| \sum_{i=1}^d \Bigl(h(x_i - z_i) - h(y_i - z_i) \Bigr) \right| \mathrm{d}z \\
&\geqslant a \int_0^1 \left| \int_{[0,1]^{d-1}} \sum_{i=1}^d \Bigl(h(x_i - z_i) - h(y_i - z_i) \Bigr) \,\mathrm{d}z_2 \cdots \mathrm{d}z_d \right| \mathrm{d}z_1 \\
&= a \int_0^1 \bigl| h(x_1 - z_1) - h(y_1 - z_1) \bigr| \,\mathrm{d}z_1 = a \|h(\cdot) - h(\cdot - x_1 + y_1)\|_{L^1[0,1]} \\
&\geqslant a c_2 |x_1 - y_1|_\circ^\alpha.
\end{aligned} \tag{3.10}$$

By symmetry, we also have the lower bound $ac|x_i - y_i|_\circ^\alpha$ for any $i \leqslant d$, and thus

$$\int_{[0,1]^d} \bigl|W(x,z) - W(y,z)\bigr| \,\mathrm{d}z \geqslant a c_2 \frac{1}{d} \sum_{i=1}^d |x_i - y_i|_\circ^\alpha \geqslant \frac{a c_2}{d} |x-y|_\circ^\alpha. \tag{3.11}$$

The estimate (3.11) implies that $\Psi_q(W) < \infty$ for every $q < d/\alpha$, similarly to (3.6). □

Proof of claim in Example 2.3. As for Example 2.2, now using Lemmas 3.2, 3.3 and 3.5. □

Remark 3.6. Although the proofs are for specific examples, the arguments suggest that the conclusions of Lemma 3.4 and 3.5, and as a consequence the conclusions of Example 2.2 and Example 2.3, are typical of graphons on $[0,1]^d$ with the given smoothness, rather than exceptional. □

References

[1] Robert J. Adler and Jonathan E. Taylor: *Random Fields and Geometry*, Springer-Verlag, Berlin, Germany, 2007.

[2] Avanti Athreya, Donniell E Fishkind, Minh Tang, Carey E. Priebe, Youngser Park, Joshua T Vogelstein, Keith Levin, Vince Lyzinski, Yichen Qin: Statistical inference on random dot product graphs: a survey. *The Journal of Machine Learning Research*, **18** (2017), no. 1, 8393–8484.

[3] Christian Borgs, Jennifer T. Chayes and László Lovász: Moments of two-variable functions and the uniqueness of graph limits. *Geom. Funct. Anal.* **19** (2010), no. 6, 1597–1619.

[4] Yasuko Chikuse: *Statistics on Special Manifolds*, Lecture Notes in Statistics, 174. Springer-Verlag, New York, 2003.

[5] Joaquim E. de Freitas, Ronaldo F. de Lima and Daniel T. dos Santos: The n-dimensional Peano curve. *São Paulo J. Math. Sci.* **13** (2019), no. 2, 678–688.

[6] Ronald A. DeVore: Nonlinear approximation. *Acta Numerica* **7** (1998), 51–150.

[7] Chao Gao, Yu Lu and Harrison H. Zhou: Rate-optimal graphon estimation. *Annals of Statistics* **43** (2015), no. 6, 2624–2652.

[8] Peter D. Hoff, Adrian E. Raftery, and Mark S. Handcock: Latent space approaches to social network analysis. *J. American Statistical Association*, **97** (2002), no. 460, 1090–1098.

[9] Svante Janson: Graphons, cut norm and distance, couplings and rearrangements. *NYJM Monographs* **4**, 2013.

[10] Eric Kolyczyk: *Statistical Analysis of Network Data*, Springer-Verlag, Berlin, Germany, 2009.

[11] László Lovász: *Large Networks and Graph Limits*, American Mathematical Society, Providence, RI, 2012.

[12] Stephen C. Milne: Peano curves and smoothness of functions. *Adv. in Math.* **35** (1980), no. 2, 129–157.

[13] Sofia C. Olhede and Patrick J. Wolfe: Network histograms and universality of blockmodel approximation. *Proceedings of the National Academy of Sciences* **111** (2014), no. 41, 14722–14727.

[14] Peter Orbanz and Daniel M. Roy: Bayesian models of graphs, arrays and other exchangeable random structures. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **37** (2014), no. 2, 437–461.

[15] Hans Sagan: *Space-filling Curves*. Springer-Verlag, New York, 1994.

[16] Antoni Zygmund: *Trigonometric Series*. 2nd ed., Cambridge University Press, 1959.

Department of Mathematics, Uppsala University, PO Box 480, SE-751 06 Uppsala, Sweden

Email address: svante.janson@math.uu.se

URL: http://www2.math.uu.se/∼svante/

Institute of Mathematics, Ecole Polytechnique Fédérale de Lausanne,Lausanne, Switzerland

Email address: sofia.olhede@epfl.ch