Submitted to the Annals of Statistics

SUPPLEMENT TO "CONSISTENT NONPARAMETRIC ESTIMATION FOR HEAVY-TAILED SPARSE GRAPHS"

By Christian Borgs[†], Jennifer T. Chayes[†], Henry Cohn[‡], and Shirshendu Ganguly[*,†]

University of California, Berkeley[†] *and Microsoft Research*[‡]

CONTENTS

APPENDIX A: COUPLINGS, METRICS, AND EQUIVALENCE

In this appendix, we expand on the material presented in Section 3.4, establishing in particular the theorems stated there. We also show that for graphons over Borel spaces, the infima in the definitions (1.1) and (3.3) of the distances δ_p and $\delta_\square$ are attained for some couplings ν, and that under the additional assumption that the underlying spaces are atomless, we have the alternative representations (3.4) for these metrics. See Lemma A.7 below for the precise statement.

We start by the following definition, whose first part just restates the definition of equivalence used in Theorem 1.1.

Definition A.1. Let W, W' be graphons over $(\Omega, \mathcal{F}, \pi)$ and $(\Omega', \mathcal{F}', \pi')$, respectively. We call W and W' *equivalent* if there exist measure-preserving maps ϕ and ϕ' from $(\Omega, \mathcal{F}, \pi)$ and $(\Omega', \mathcal{F}', \pi')$ to a third probability space $(\Omega'', \mathcal{F}'', \pi'')$ and a graphon U on $(\Omega'', \mathcal{F}'', \pi'')$ such that $W = U^\phi$ and $W' = U^{\phi'}$ almost everywhere. We call W and W' *isomorphic modulo* 0 if there exists a map $\phi\colon \Omega \to \Omega'$ such that ϕ is an isomorphism aside from sets of measure zero and $W = (W')^\phi$ almost everywhere.

[*]Supported by an internship at Microsoft Research New England.

Remark A.2. Our notion of equivalence is closely related to the notion of "weak isomorphism" from [3], the only difference being that in [3] the maps ϕ and ϕ' were required to be measure preserving with respect to the completions of the spaces $(\Omega, \mathcal{F}, \pi)$ and $(\Omega', \mathcal{F}', \pi')$. We will not use the term weak isomorphism since we want to avoid the impression that it implies that the underlying probability spaces are isomorphic after removing suitable sets of measure zero. It does not; see Example A.3 for a counterexample.

Example A.3 also shows that for two graphons to be equivalent we in general do need two maps ϕ and ϕ'; i.e., it shows that in general, two equivalent graphons cannot be obtained from each other by a single pullback.

Example A.3. Let $\Omega = [4]$ and $\Omega' = [6]$, both equipped with the uniform distribution. Define W and W' to be 0 if both arguments are even or both arguments are odd, and set both of them to 1 otherwise. It is easy to see that they are equivalent: indeed, let $\Omega'' = \{1, 2\}$ and define $\phi\colon [4] \to [2]$ and $\psi\colon [6] \to [2]$ by mapping even elements to 2 and odd elements to 1. Setting U to 1 if its two arguments are different and to 0 otherwise, we see that $W = U^\phi$ and $W' = U^\psi$. This shows that in general, we cannot restrict ourselves to a single, measure-preserving map $\phi\colon \Omega \to \Omega'$, since there is simply no measure-preserving map between Ω and Ω'.

But even if both probability spaces are $[0, 1]$ equipped with the uniform measure (in which case there are many measure-preserving maps between the two), we can in general not find a measure-preserving map such that $W' = W^\phi$ or the other way around. To see this, let $\phi_k(x) = kx \bmod 1$, define $W_1(x, y) = xy$, and let $W_k = W_1^{\phi_k}$. Then there is no measure-preserving transformation $\phi\colon [0, 1] \to [0, 1]$ such that $W_2^\phi = W_3$ or $W_3^\phi = W_2$; see Example 8.2 in [7] for the proof.

There is however, a special case where it is possible to just use a single map, namely the case where W and W' are twin-free Borel graphons. Here a graphon is called a *Borel graphon* if the underlying probability space is a *Borel space*, i.e., a space that is isomorphic to a Borel subset of a complete separable metric space equipped with an arbitrary probability measure with respect to the Borel σ-algebra. A graphon W is called *twin-free* if the set of twins of W has measure zero, where a *twin* is a point x in the underlying probability space for which there is another point y such that $W(x, \cdot)$ is equal to $W(y, \cdot)$ almost everywhere. Note that in Example A.3 above, the graphons U and W_1 are twin-free, while W, W', and W_k for $k \geq 2$ are not.

THEOREM A.4. *Let W and W' be twin-free Borel graphons. Then W and W' are equivalent if and only if they are isomorphic modulo* 0.

The theorem can easily be deduced from the results of [3], and is proved below.

The next theorem is an extension of Theorem 3.5. To state it, we define a *standard Borel graphon* to be a graphon over a probability space that is the disjoint union of an interval $[0,p]$ equipped with the uniform distribution and the usual Borel σ-algebra, plus a countable number of isolated points $\{x_j\}_{j\in J}$ with nonzero mass p_j for each of them, allowing for the special cases where either the set of atoms or the interval $[0,p]$ is absent. The former is the case of graphons over $[0,1]$, while the latter is the case of block models over $[k]$ equipped with a probability measure in Δ_k.

THEOREM A.5. *Let W be a graphon over an arbitrary probability space $(\Omega,\mathcal{F},\pi)$.*

(i) There exists an equivalent graphon over $[0,1]$ *equipped with the uniform distribution.*

(ii) There exists a twin-free standard Borel graphon U and a measure-preserving map ϕ from $(\Omega,\mathcal{F},\pi)$ to the space on which U is defined such that $W=U^\phi$ almost everywhere, showing in particular that W is equivalent to a twin-free standard Borel graphon.

REMARK A.6. The above theorem states that for any graphon W, we can find both an equivalent graphon U over $[0,1]$ and an equivalent twin-free standard Borel graphon $\tilde{U}$. But in general, it is not possible to find a single equivalent graphon U which is both twin-free and a graphon over $[0,1]$, as the example of a block model shows, since any representation of it over $[0,1]$ has uncountably many twins.

Next, we give a precise formulation of Remark 3.1(i).

LEMMA A.7. *Let $p\geq 1$ and let W and W' be L^p graphons over two Borel spaces $(\Omega,\mathcal{F},\pi)$ and $(\Omega',\mathcal{F}',\pi')$. Then the following hold:*

(i) The infima in (1.1) *and* (3.3) *are attained for some couplings ν.*

(ii) If $(\Omega,\mathcal{F},\pi)$ and $(\Omega',\mathcal{F}',\pi')$ are atomless, then the distances $\delta_p(W,W')$ and $\delta_\square(W,W')$ can be expressed as

$$\delta_p(W,W')=\inf_\phi\|W-(W')^\phi\|_p=\inf_\Phi\|W-(W')^\Phi\|_p$$

and

$$\delta_\square(W,W')=\inf_\phi\|W-(W')^\phi\|_\square=\inf_\Phi\|W-(W')^\Phi\|_\square,$$

where the infima over ϕ are over measure-preserving maps from Ω to Ω' and the infima over Φ are over isomorphisms from Ω to Ω'.

For the cut metric, the first statement is a special case of Theorem 6.16 in [7] (see also Lemma 2.6 in [1], which proves the statement for bounded graphons over $[0,1]$), while the second is essentially given in Lemma 3.5 in [4]. Specifically, while Lemma 3.5 in [4] was only stated for bounded graphons over $[0,1]$, the generalization to unbounded graphons over an atomless Borel space is straightforward. The proofs for the distance δ_p are virtually identical. For the convenience of the reader, we sketch them below.

Note that the first statement does not hold without the assumption that $(\Omega,\mathcal{F},\pi)$ and $(\Omega',\mathcal{F}',\pi')$ are Borel spaces; see, for example, Example 8.13 in [7] for a counterexample. Similarly, the assumption that $(\Omega,\mathcal{F},\pi)$ and $(\Omega',\mathcal{F}',\pi')$ are atomless is needed for the second statement to hold; see Remark 6.10 in [7]. (Indeed, the condition involving Φ does not even make sense unless Ω and Ω' are isomorphic, but all atomless Borel spaces are isomorphic by Theorem A.7 in [7]. For arbitrary probability spaces there may not even be any measure-preserving maps from Ω to Ω'.)

Proof. We begin with part (i). For the cut metric, this is a special case of Theorem 6.16 in [7]. The proof for the metric δ_p is very similar. For the convenience of the reader, we give the proof below, combining proof techniques from [7] and [1].

Let $\mathcal{M}$ be the set of all probability measures on $\Omega\times\Omega'$ for which the marginals are π and π'. We first observe that $\mathcal{M}$ is compact in the weak* topology. To see why, first note that by Theorem A.4(iv) in [7], the measurable spaces $(\Omega,\mathcal{F})$ and $(\Omega',\mathcal{F}')$ are either countable (with all subsets measurable) or isomorphic to $[0,1]$ with the Borel σ-algebra. Let $\mathcal{A}_0$ be the set of all $A\subseteq\Omega\times\Omega'$ that are products of intervals with rational endpoints in the $[0,1]$ case and finite sets in the countable case. Since $\mathcal{A}_0$ is countable, any sequence of measures $\nu_n\in\mathcal{M}$ has a subsequence ν'_n such that $\nu'_n(A)$ converges for all $A\in\mathcal{A}_0$. Since $\mathcal{A}_0$ generates the product σ-algebra on $\Omega\times\Omega'$, the limit can be extended to a probability measure μ on $\Omega\times\Omega'$, which can easily be checked to have π and π' as marginals, implying that $\mu\in\mathcal{M}$.

Consider a sequence of couplings ν_n such that

$$\begin{aligned}&\delta_p(W,W') =\\ &\lim_{n\to\infty}\left(\int\Big|W(x,y)-W'(x',y')\Big|^p\,d\nu_n(x,x')\,d\nu_n(y,y')\right)^{1/p}.\end{aligned}\tag{A.1}$$

By the compactness of $\mathcal{M}$, we may pass to a subsequence (which we again denote by ν_n) for which there is a limit $\nu \in \mathcal{M}$ such that $\nu_n(A) \to \nu(A)$ for all $A \in \mathcal{A}_0$. Since $\nu \in \mathcal{M}$,

$$\delta_p(W, W') \leq \left(\int \Big| W(x,y) - W'(x',y') \Big|^p \, d\nu(x,x') \, d\nu(y,y') \right)^{1/p}.$$

To prove a matching lower bound we fix $\varepsilon > 0$ to be sent to zero later. By (A.1), we can find an n_0 such that

$$\delta_p(W, W') \geq \left(\int \Big| W(x,y) - W'(x',y') \Big|^p \, d\nu_n(x,x') \, d\nu_n(y,y') \right)^{1/p} - \varepsilon.$$

for all $n \geq n_0$. Since $W \in L^p$, we can find an M such that $\|W 1_{W \geq M}\|_p \leq \varepsilon$, and since $W 1_{W<M}$ is bounded, we can find a graphon $\tilde{W}$ which is a finite sum of the form $\tilde{W} = \sum_{i,j} \beta_{ij} 1_{A_i \times A_j}$ with $A_i \in \mathcal{A}_0$ such that $\|W 1_{W<M} - \tilde{W}\|_p \leq \varepsilon$, implying in particular $\|W - \tilde{W}\|_p \leq 2\varepsilon$. In a similar way, we can find $\tilde{W}'$ of the form $\tilde{W}' = \sum_{k,\ell} \beta'_{k\ell} 1_{B_k \times B_\ell}$ with $B_i \in \mathcal{A}_0$ and $\|W' - \tilde{W}'\|_p \leq 2\varepsilon$. As a consequence

$$\begin{aligned}
\delta_p(W, W') &\geq \left(\int \Big| \tilde{W}(x,y) - \tilde{W}'(x',y') \Big|^p \, d\nu_n(x,x') \, d\nu_n(y,y') \right)^{1/p} - 5\varepsilon \\
&= \left(\sum_{i,j,k,\ell} |\beta_{ij} - \beta'_{k\ell}|^p \nu_n(A_i \times B_k) \, \nu_n(A_j \times B_\ell) \right)^{1/p} - 5\varepsilon
\end{aligned}$$

for all $n \geq n_0$. We can take the limit as $n \to \infty$ on the right side, to obtain the bound

$$\begin{aligned}
\delta_p(W, W') &\geq \left(\sum_{i,j,k,\ell} |\beta_{ij} - \beta'_{k\ell}|^p \nu(A_i \times B_k) \, \nu(A_j \times B_\ell) \right)^{1/p} - 5\varepsilon \\
&= \left(\int \Big| \tilde{W}(x,y) - \tilde{W}'(x',y') \Big|^p \, d\nu(x,x') \, d\nu(y,y') \right)^{1/p} - 5\varepsilon \\
&\geq \left(\int \Big| W(x,y) - W'(x',y') \Big|^p \, d\nu(x,x') \, d\nu(y,y') \right)^{1/p} - 9\varepsilon.
\end{aligned}$$

Since ε was arbitrary, this proves part (i) of the lemma.

We now turn to part (ii). All atomless Borel spaces are isomorphic to $[0,1]$ (with the Borel σ-algebra and uniform distribution), by Theorem A.7 in [7]. Thus, we can assume without loss of generality that Ω and Ω' are both $[0,1]$.

Choosing z uniformly at random from $[0,1]$, the map $z \mapsto (z, \phi(z))$ provides a coupling showing that $\delta_p(W,W') \le \inf_\phi \|W - (W')^\phi\|_p$ and $\delta_\square(W,W') \le \inf_\phi \|W - (W')^\phi\|_\square$. It is also obvious that $\inf_\phi \|W - (W')^\phi\|_p \le \inf_\Phi \|W - (W')^\Phi\|_p$ and $\inf_\phi \|W - (W')^\phi\|_\square \le \inf_\Phi \|W - (W')^\Phi\|_\square$.

To prove equality, one first approximates W and W' by piecewise constant functions (more precisely, graphons on $[n]$ equipped with the uniform measure), and then approximates an arbitrary coupling of two uniform measures on $[n]$ by a bijection on a "blow-up" $[nk]$ of $[n]$. Mapping this bijection back to an isomorphism $\Phi\colon [0,1] \to [0,1]$ then gives a lower bound on $\delta_p(W,W')$ in terms of $\inf_\Phi \|W^\Phi - W'\|_p$, minus some error which can be taken to be arbitrarily small. The details are very similar to the proof of Lemma 3.5 in [4], which proves equality for the cut norm when W and W' are bounded, and we leave them to the reader. Note that the generalization to unbounded graphons is straightforward, given that $\|W 1_{W \ge M}\|_p \to 0$ as $M \to \infty$ and $\|W 1_{W\ge M}\|_\square \le \|W 1_{W \ge M}\|_1$. $\square$

In the remainder of this appendix, we prove Theorem 3.6 from Section 3.4, as well as Theorem A.4 and Theorem A.5 (which encompasses Theorem 3.5). We rely heavily on both the results and the techniques of [3] and [7]; see also [1]. Before turning to these proofs, we relate the notion of equivalence from Definition A.1 to the notion of "weak isomorphism" from [3], which requires the maps ϕ and ϕ' to be measure preserving with respect to the completion of the spaces $(\Omega, \mathcal{F}, \pi)$ and $(\Omega', \mathcal{F}', \pi')$. It is clear that equivalence implies weak isomorphism, since maps that are measurable with respect to $(\Omega, \mathcal{F}, \pi)$ and $(\Omega', \mathcal{F}', \pi')$ are clearly measurable with respect to their completions. We can also turn this around, at least when the third space is a *Lebesgue space*, i.e., the completion of a Borel space. This follows from part (i) of the following technical lemma.

LEMMA A.8. *Let W and W' be graphons over two probability spaces $(\Omega, \mathcal{F}, \pi)$ and $(\Omega', \mathcal{F}', \pi')$, respectively.*

(i) Assume that there exist measure-preserving maps ϕ and ϕ' from the completions of $(\Omega, \mathcal{F}, \pi)$ and $(\Omega', \mathcal{F}', \pi')$ to a Lebesgue space $(\Omega'', \mathcal{F}'', \pi'')$ and a graphon U over $(\Omega'', \mathcal{F}'', \pi'')$ such that $W = U^\phi$ and $W' = U^{\phi'}$ almost everywhere. Then there exists a standard Borel graphon $\tilde U$ and measure-preserving maps $\tilde\phi$ and $\tilde\phi'$ from $(\Omega, \mathcal{F}, \pi)$ and $(\Omega', \mathcal{F}', \pi')$ to the Borel space on which $\tilde U$ is defined such that $W = \tilde U^{\tilde\phi}$ and $W' = \tilde U^{\tilde\phi'}$ almost everywhere. If U is twin-free, then $\tilde U$ can be chosen to be twin-free as well.

(ii) If $(\Omega, \mathcal{F}, \pi)$ and $(\Omega', \mathcal{F}', \pi')$ are Borel spaces and W and W' are isomorphic modulo 0 when considered as graphons over the completion of

$(\Omega, \mathcal{F}, \pi)$ and $(\Omega', \mathcal{F}', \pi')$, then they are also isomorphic modulo 0 *as graphons over $(\Omega, \mathcal{F}, \pi)$ and $(\Omega', \mathcal{F}', \pi')$.*

PROOF. (i) Since every Lebesgue space is isomorphic modulo 0 to the union of an interval $[0, p]$ and a collection of atoms x_i (see Theorem A.10 in [7]), we may without loss of generality assume that $(\Omega'', \mathcal{F}'', \pi'')$ is of this form. Assume without loss of generality that the atoms are represented as points $x_i \in (p, 1]$, so that ϕ takes values in $[0, 1]$. Noting that $\mathcal{F}''$ is the completion of a Borel σ-algebra $\mathcal{B}''$, define $\tilde{U}$ as the conditional expectation $\mathbb{E}[U \mid \mathcal{B}'' \times \mathcal{B}'']$. Then $\tilde{U}$ is a Borel graphon such that $U = \tilde{U}$ almost everywhere. Since ϕ is measure preserving from the completion $(\Omega, \bar{\mathcal{F}}, \pi)$ of $(\Omega, \mathcal{F}, \pi)$ to $(\Omega'', \mathcal{F}'', \pi'')$, it is also measure preserving from $(\Omega, \bar{\mathcal{F}}, \pi)$ to $(\Omega'', \mathcal{B}'', \pi'')$. Replacing ϕ by the conditional expectation $\tilde{\phi} = \mathbb{E}[\phi \mid \mathcal{F}]$, we obtain a measure-preserving map $\tilde{\phi}$ from $(\Omega, \mathcal{F}, \pi)$ to $(\Omega'', \mathcal{B}'', \pi'')$ such that $W = \tilde{U}^{\tilde{\phi}}$ almost everywhere. If U is twin-free, then so is $\tilde{U}$.

(ii) The completions of $(\Omega, \mathcal{F}, \pi)$ and $(\Omega', \mathcal{F}', \pi')$ are Lebesgue spaces. Since every Lebesgue space is isomorphic modulo 0 to the disjoint union of an interval $[0, p]$ (equipped with the Lebesgue σ-algebra and the uniform measure) and countably many atoms x_i, we have that as graphons over the completion of $(\Omega, \mathcal{F}, \pi)$ and $(\Omega', \mathcal{F}', \pi')$, both W and W' are isomorphic modulo 0 to a graphon U over such a space. Proceeding as in the proof of (i), we can then replace U by a Borel graphon $\tilde{U}$ such that W and W' are isomorphic modulo 0 to the graphon $\tilde{U}$, which in particular implies that W and W' are isomorphic modulo 0. $\square$

PROOF OF THEOREM A.4. If W and W' are isomorphic modulo 0, they are clearly equivalent. Assume on the other hand that W and W' are equivalent. Moving from $(\Omega, \mathcal{F}, \pi)$ and $(\Omega', \mathcal{F}', \pi')$ to their completion, we obtain graphons which are defined on a Lebesgue space and are weakly isomorphic in the sense of [3]. For bounded graphons, we can then use Theorem 2.1 of [3] to conclude that W and W' are isomorphic modulo 0 as graphons over the completion of $(\Omega, \mathcal{F}, \pi)$ and $(\Omega', \mathcal{F}', \pi')$. By Lemma A.8, this implies that they are also isomorphic modulo 0 as graphons over $(\Omega, \mathcal{F}, \pi)$ and $(\Omega', \mathcal{F}', \pi')$.

If W and W' are unbounded, let $\widetilde{W} = \tanh W$ and $\widetilde{W}' = \tanh W'$. Clearly, W and W' are equivalent if and only if $\widetilde{W}$ and $\widetilde{W}'$ are equivalent, and W and W' are isomorphic modulo 0 if and only if $\widetilde{W}$ and $\widetilde{W}'$ are isomorphic modulo 0. Therefore the unbounded case follows from the bounded case. $\square$

PROOF OF THEOREM A.5. For bounded graphons, the analogous statement for graphons over a Lebesgue space was proved in [3]; in particular, by Corollary 3.3 from [3], we can find a twin-free graphon U over a Lebesgue space $(\Omega', \mathcal{F}', \pi')$ and a measure-preserving map ϕ from the completion of $(\Omega, \mathcal{F}, \pi)$ to $(\Omega', \mathcal{F}', \pi')$ such that $W = U^{\phi}$ almost everywhere. By Lemma A.8, this implies the existence of a twin-free standard Borel graphon $\tilde{U}$ on a Borel space $(\tilde{\Omega}, \tilde{\mathcal{F}}, \tilde{\pi})$ and a measure-preserving map from $(\Omega, \mathcal{F}, \pi)$ to $(\tilde{\Omega}, \tilde{\mathcal{F}}, \tilde{\pi})$ such that $W = \tilde{U}^{\tilde{\phi}}$ almost everywhere, which proves (ii) for bounded graphons. Statement (i) follows from (ii) by expanding the atoms x_i in $\tilde{\Omega}$ into intervals of widths $p_i = \tilde{\pi}(x_i)$.

To reduce the case of unbounded graphons to the case of bounded graphons, we again use the transformation $W \mapsto \tanh W$, which maps arbitrary graphons to bounded graphons. □

Proof of Theorem 3.6. We first note that the implications (iii) ⇒ (ii) ⇒ (i) are trivial. So all that remains to prove is that (i) ⇒ (iii), and by Theorem 3.5, it will be enough to prove this for graphons W and W' over $[0,1]$ equipped with the uniform distribution.

Assume thus that W and W' are graphons over $[0,1]$ with $\delta_\square(W, W') = 0$. By Lemma A.7 this implies that W and W' can be coupled in such a way that $\|W - W'\|_\square = 0$, which in turn implies that $W(x,y) = W'(x',y')$ almost surely with respect to this coupling. As a consequence, $\delta_\square(\tanh W, \tanh W') = 0$. By the results of [3], this implies that $\tanh W$ and $\tanh W'$ are equivalent, which in turn shows that W and W' are equivalent, as required. □

APPENDIX B: ESTIMATING DEGREE DISTRIBUTIONS

In this appendix, we show that a good approximation of a graphon in the cut metric implies a good approximation for its degree distribution. We define the *normalized degree* of a vertex $x \in V(G)$ as $d_x/\bar{d}$, where d_x is its degree and $\bar{d} = \frac{1}{|V(G)|}\sum_{x \in V(G)} d_x$ is the average degree. The *normalized degree distribution* of G is the empirical distribution of the normalized degrees, with cumulative distribution function

$$D_G(\lambda) = \frac{1}{|V(G)|}\sum_{x \in V(G)} 1_{d_x \le \lambda \bar{d}}.$$

In a similar way, we define the *degrees* of a normalized graphon $W\colon \Omega \times \Omega \to [0,\infty)$ as the random variable

$$W_x = \int_\Omega W(x,y)\, d\pi(y),$$

where x is chosen according to the probability measure π on Ω. This random variable has cumulative distribution function

$$D_W(\lambda) = \pi(\{x : W_x \le \lambda\}).$$

Recalling that convergence in distribution can be formulated as convergence in the Lévy-Prokhorov distance, we say that the normalized degree distributions of a sequence G_n of graphs converge to the degree distribution of W if $d_{\mathrm{LP}}(D_{G_n}, D_W) \to 0$, where as usual, the Lévy-Prokhorov distance d_{LP} between two distribution functions D and D' is defined by

$$d_{\mathrm{LP}}(D, D') = \inf\{\varepsilon > 0 : D'(\lambda-\varepsilon)-\varepsilon \leq D(\lambda) \leq D'(\lambda+\varepsilon)+\varepsilon \text{ for all } \lambda \in \mathbb{R}\}.$$

Our next theorem implies that convergence in the cut metric implies convergence of the normalized degree distributions. Combined with Theorem 3.10, this shows that a.s., the normalized degree distributions of a sequence of W-random graphs converge to the degree distribution of W as long as $n\rho_n \to \infty$ and $\rho_n \to 0$. Indeed, observing that for any graph G, the normalized degree distribution D_G is equal to the degree distribution of $\frac{1}{\|A(G)\|_1}A(G)$ considered as a graphon over $V(G)$ equipped with the uniform distribution, both statements follow immediately from the following theorem.

THEOREM B.1. *Let U and W be two normalized graphons. Then*

$$d_{\mathrm{LP}}(D_U, D_W) \leq \sqrt{2\delta_{\square}(U, W)}.$$

The proof of Theorem B.1 will make use of the following lemma.

LEMMA B.2. *Let U and W be two normalized graphons over the same probability space Ω. If x is chosen at random from Ω, then*

$$\Pr(|W_x - U_x| \geq \varepsilon) \leq \frac{2}{\varepsilon}\|U - W\|_{\square}.$$

PROOF. We have

$$\begin{aligned}\Pr(|W_x - U_x| \geq \varepsilon) &\leq \frac{1}{\varepsilon}\mathbb{E}[|W_x - U_x|] \\ &= \frac{1}{\varepsilon}\mathbb{E}[(W_x - U_x)1_{W_x \geq U_x}] + \frac{1}{\varepsilon}\mathbb{E}[(U_x - W_x)1_{W_x \leq U_x}].\end{aligned}$$

Defining S as the set of points $x \in \Omega$ such that $W_x \geq U_x$ and $\tilde{S}$ as the set of points $x \in \Omega$ such that $W_x \leq U_x$, we write the right side as

$$\frac{1}{\varepsilon}\int_{[0,1]\times S}(W - U) + \frac{1}{\varepsilon}\int_{[0,1]\times \tilde{S}}(U - W) \leq \frac{2}{\varepsilon}\|U - W\|_{\square},$$

as desired. ∎

Proof of Theorem B.1. To prove the theorem, we will prove that for two arbitrary graphons and all $\lambda \in \mathbb{R}$ and $\varepsilon > 0$,

$$D_W(\lambda) \le D_U(\lambda + \varepsilon) + 2\frac{\delta_\square(U, W)}{\varepsilon}. \tag{B.1}$$

Because the degree distributions of equivalent graphons are identical, it will be enough to prove (B.1) for two graphons over $[0, 1]$, with an upper bound of $\|U - W\|_\square$ instead of $\delta_\square(U, W)$.

To this end, we estimate the probability that U_x and W_x differ by at least ε by Lemma B.2. As a consequence,

$$\begin{aligned} D_W(\lambda) &= \Pr[W_x \le \lambda] \\ &\le \Pr[U_x \le \lambda + \varepsilon] + \Pr[|U_x - W_x| \ge \varepsilon] \\ &\le D_U(\lambda + \varepsilon) + \frac{2}{\varepsilon}\|U - W\|_\square, \end{aligned}$$

which proves (B.1) and hence the theorem. □

APPENDIX C: AUXILIARY RESULTS

In this appendix, we prove several auxiliary results needed for various proofs in this paper.

C.1. Existence of approximating block models. Our first set of results concern the approximation of an arbitrary graphon by a stochastic block model. Recall the definition of $\mathcal{B}_{\ge\kappa}$ as the set of all block models with minimal block size at least κ,

$$\mathcal{B}_{\ge\kappa} = \{(\mathbf{p}, B) \in \mathcal{B} : \min_i p_i \ge \kappa\}.$$

For an arbitrary probability space $(\Omega, \mathcal{F}, \mu)$ and integrable function $U \colon \Omega \times \Omega \to \mathbb{R}$, we define $H_n(U)$ to be the $n \times n$ matrix with entries

$$(H_n(U))_{ij} = U(x_i, x_j) \quad \text{for} \quad i \ne j \tag{C.1}$$

and $(H_n(U))_{ii} = 0$; in contrast to the definitions of $G_n(W)$ and $Q_n(W)$, which we will only use for graphons, the latter notation will be used even if U takes values in $\mathbb{R}$, rather than in $[0, \infty)$.

Lemma C.1. *Let $1 \le p < \infty$, let W be an L^p graphon, and let $\varepsilon^{(p)}_{\ge\kappa}(W)$ be as in (2.1). Then the infimum in (2.1) is achieved for some $W' \in \mathcal{B}_{\ge\kappa}$ that has norm $\|W'\|_p \le 2\|W\|_p$. Furthermore, $\varepsilon^{(p)}_{\ge\kappa}(W) \to 0$ as $\kappa \to 0$.*

PROOF. We clearly have $\varepsilon^{(p)}_{\geq\kappa}(W) = \inf_{W'} \delta_p(W, W') \leq \|W\|_p$, so by the triangle inequality, we only need to consider block models W' with $\|W'\|_p \leq 2\|W\|_p$. Again by the triangle inequality, the distance $\delta_p(W, W')$ is continuous in W', which implies that the infimum is actually a minimum.

To see that $\varepsilon^{(p)}_{\geq\kappa}(W) \to 0$ as $\kappa \to 0$, we first replace W by an equivalent graphon U over $[0, 1]$, and then use the approximation U_n to U given by averaging over the partition consisting of consecutive intervals of length $1/n$. This approximation is a block model with minimal block size $1/n$, and it is not hard to prove that it converges to U. (For example, because continuous functions are dense in L^p, we can approximate U with a continuous function. Then convergence follows from uniform continuity and the fact that averaging over a partition is a contraction in L^p.) □

When applying the lemma, we will sometimes be constrained to use only block models whose block sizes are all a multiple of $1/n$, i.e., block models in

$$\mathcal{B}_{\geq\kappa,n} = \{(\mathbf{p}, B) \in \mathcal{B} : \text{for all } i,\ p_i n \in \mathbb{Z} \text{ and } p_i n \geq \lfloor n\kappa \rfloor\}.$$

Note that $\mathcal{B}_{\geq\kappa,n}$ naturally corresponds to the set $\mathcal{A}_{n,\geq\kappa}$ of $n \times n$ block matrices A such that each block in A has size at least $\lfloor n\kappa \rfloor$, via

$$\{\mathsf{W}[A] : A \in \mathcal{A}_{n,\geq\kappa}\} = \{\mathsf{W}[\mathbf{p}, B] : (\mathbf{p}, B) \in \mathcal{B}_{\geq\kappa,n}\}. \tag{C.2}$$

Our next lemma shows that every block model in $\mathcal{B}_{\geq\kappa}$ can be well approximated by a block model in $\mathcal{B}_{\geq\kappa,n}$, and it also shows that $\varepsilon^{(p)}_{\geq\kappa}$ can be bounded from above in terms of a minimum over $\mathcal{B}_{\geq\kappa,n}$.

LEMMA C.2. *Let $\kappa \in (0, 1]$. Then there exists a constant $n_0(\kappa)$ such that for all $p \geq 1$, the following holds:*

If $W' \in \mathcal{B}_{\geq\kappa}$ is a block model on $[k]$, then the labels in $[k]$ can be reordered in such a way that for each $n \geq 1/\kappa$ there exists a block model $W'' \in \mathcal{A}_{n,\geq\kappa}$ with

$$\hat{\delta}_p(W'', \mathsf{W}[W']) \leq 2\sqrt[p]{\frac{2}{\kappa n}}\|W'\|_p$$

(recall that $\hat{\delta}_p$ is defined in (3.1)). For all L^p graphons W, if $n \geq n_0(\kappa)$, then

$$\varepsilon^{(p)}_{\geq\kappa}(W) \leq \min_{W'' \in \mathcal{B}_{\geq\kappa,n}} \delta_p(W'', W) + 4\sqrt[p]{\frac{2}{\kappa^2 n}}\|W\|_p.$$

The proof of the lemma relies on two other lemmas, the first of which is an easy consequence of the law of large numbers for U-statistics.

LEMMA C.3. *Let $(\Omega, \mathcal{F}, \pi)$ be a probability space, and let $W \colon \Omega \times \Omega \to \mathbb{R}$ be in L^p for some $p \geq 1$. Then $\|H_n(W)\|_p \to \|W\|_p$ a.s.*

PROOF. Define $U = |W|^p$, and choose $x_1, \dots, x_n$ i.i.d. with distribution π. Then

$$\|H_n(W)\|_p^p = \frac{1}{n^2} \sum_{i \neq j} |W(x_i, x_j)|^p = \frac{1}{n^2} \sum_{i \neq j} U(x_i, x_j).$$

By the strong law of large numbers for U-statistics (see, for example, [6]), the right side converges to $\|U\|_1 = \|W\|_p^p$ as claimed. □

Next we state the second lemma needed to prove Lemma C.2. We use λ to denote the Lebesgue measure on $[0,1]$ or $[0,1]^2$ (depending on the context), and, as usual, we use $A \triangle B$ to denote the symmetric difference of two sets A, B, i.e., $A \triangle B = (A \setminus B) \cup (B \setminus A)$.

LEMMA C.4. *Let W and W' be of the form $W = \sum_{i,j} B_{ij} 1_{Y_i \times Y_j}$ and $W' = \sum_{i,j} B_{ij} 1_{Y'_i \times Y'_j}$, where B is a $k \times k$ matrix, and $(Y_1, \dots, Y_k)$ and $(Y'_1, \dots, Y'_k)$ are partitions of $[0,1]$. If $\lambda(Y_i \triangle Y'_i) \leq \varepsilon \lambda(Y_i)$ for all i, then*

$$\|W - W'\|_p \leq 2 \sqrt[p]{\varepsilon(1+\varepsilon)} \|W\|_p.$$

PROOF. We begin by writing

$$\begin{aligned} W - W' &= \sum_{i,j} B_{ij} (1_{Y_i \times Y_j} - 1_{Y'_i \times Y'_j}) \\ &= \sum_{i,j} B_{ij} (1_{(Y_i \times Y_j) \setminus (Y'_i \times Y'_j)} - 1_{(Y'_i \times Y'_j) \setminus (Y_i \times Y_j)}). \end{aligned}$$

Each point of $[0,1]^2$ is in at most two of the sets $(Y_i \times Y_j) \setminus (Y'_i \times Y'_j)$ and $(Y'_i \times Y'_j) \setminus (Y_i \times Y_j)$, namely one of each of these two types, and we can therefore apply the inequality $|\sum_\ell x_\ell|^p \leq 2^{p-1} \sum_\ell |x_\ell|^p$, which holds by Hölder's inequality whenever at most two of the summands x_ℓ are nonzero. Thus,

$$\|W - W'\|_p^p \leq 2^{p-1} \sum_{i,j} |B_{ij}|^p \lambda\big((Y_i \times Y_j) \triangle (Y'_i \times Y'_j)\big).$$

We have

$$(Y_i \times Y_j) \triangle (Y'_i \times Y'_j) \subseteq \big((Y_i \cup Y'_i) \times (Y_j \triangle Y'_j)\big) \cup \big((Y_i \triangle Y'_i) \times (Y_j \cup Y'_j)\big).$$

Combining this containment with $\lambda(Y_i \triangle Y_i') \leq \varepsilon\lambda(Y_i)$ and $\lambda(Y_i \cup Y_i') \leq (1+\varepsilon)\lambda(Y_i)$ yields

$$\|W - W'\|_p^p \leq 2^p\varepsilon(1+\varepsilon)\sum_{i,j}|B_{ij}|^p\lambda(Y_i \times Y_j) = 2^p\varepsilon(1+\varepsilon)\|W\|_p^p,$$

as desired. □

REMARK C.5. A slight variation of the above proof also shows that

$$\|W - W'\|_p \leq \max_i \frac{2}{\lambda(Y_i)^{2/p}}\|W\|_p,$$

no matter how large the measure of the symmetric differences $Y_i \triangle Y_i'$ is. To see this, just bound

$$\begin{aligned}\|W - W'\|_p^p &\leq 2^{p-1}\sum_{i,j}|B_{ij}|^p\big(\lambda(Y_i \times Y_j) + \lambda(Y_i' \times Y_j')\big)\\ &\leq 2^p\|W\|_p^p \max_i \left(\frac{\lambda(Y_i')}{\lambda(Y_i)}\right)^2\\ &\leq 2^p\|W\|_p^p \max_i \frac{1}{\lambda(Y_i)^2}.\end{aligned}$$

PROOF OF LEMMA C.2. If $\kappa = 1$, $\mathcal{B}_{\geq\kappa,n}(W) = \mathcal{B}_{\geq\kappa}(W)$ and there is nothing to prove. We may therefore assume without loss of generality that $\kappa \in (0,1)$.

To prove the first bound, we write W' as $(\mathbf{p}, B)$ and reorder the elements of $[k]$ so that $p_1 \leq p_2 \leq \cdots \leq p_k$. Also, without loss of generality, we may remove all labels with $p_i = 0$, so that $p_i \geq \kappa$ for all i. Define $W'' = (\mathbf{p}'', B)$, where $\mathbf{p}''$ is obtained from $\mathbf{p}$ so that for each i, $p_1''+\cdots+p_i''$ equals $p_1+\cdots+p_i$ rounded to the nearest multiple of $1/n$ (with the convention that in the case of ties, we choose the point to the left). After embedding both W' and W'' into the space of graphons on $[0,1]$, we can write the resulting graphons $\widetilde{W''} = \mathsf{W}[W'']$ and $\widetilde{W'} = \mathsf{W}[W']$ in the form $\widetilde{W''} = \sum_{i,j} B_{ij}1_{Y_i''\times Y_j''}$ and $\widetilde{W'} = \sum_{i,j} B_{ij}1_{Y_i \times Y_j}$, where Y_i and Y_i'' are intervals whose endpoints differ by at most $1/(2n)$. As a consequence $\lambda(Y_i \triangle Y_i'') \leq \frac{1}{n} \leq \frac{1}{\kappa n}\lambda(Y_i)$. By Lemma C.4 and the fact that $\frac{1}{\kappa n} \leq 1$, this implies that

$$\|\mathsf{W}[W'] - \mathsf{W}[W'']\|_p^p \leq \frac{2^{p+1}}{\kappa n}\|W'\|_p^p. \tag{C.3}$$

To complete the proof of the first bound, all we need to show is that $W'' \in \mathcal{B}_{\geq\kappa,n}$, which means we need to show that $np_i'' = n\lambda(Y_i'') \geq \lfloor\kappa n\rfloor$ for all i.

Let i_0 be the first i such that np_i is not an integer. For $i < i_0$, we then have $np''_i = np_i \geq \kappa n \geq \lfloor \kappa n \rfloor$. On the other hand, for $i \geq i_0$, we can use $|np_i - np''_i| \leq 1$, which follows from $|n(p_1 + \cdots + p_i) - n(p''_1 + \cdots + p''_i)| \leq 1/2$. We then conclude that $np''_i \geq np_i - 1 \geq np_{i_0} - 1 > \lfloor np_{i_0} \rfloor - 1 \geq \lfloor \kappa n \rfloor - 1$, where we used that np_{i_0} is not an integer. Since np''_i is an integer, this implies $np''_i \geq \lfloor n\kappa \rfloor$, which shows that $W'' \in \mathcal{B}_{\geq \kappa, n}$. Identifying W'' with the corresponding matrix in $\mathcal{A}_{n,\geq \kappa}$, this proves the first bound.

To prove the second bound we first observe that the minimizer $W'' = (\mathbf{p}'', B) \in \mathcal{B}_{\geq \kappa, n}$ obeys the bound $\|W''\|_p \leq 2\|W\|_p$. Our task is now to find a block model $W' \in \mathcal{B}_{\geq \kappa}$ that approximates W'' in the norm δ_p. Let k'' be the number of classes in W''; again, we assume without loss of generality that they are all non-empty, which means we have that $p''_i \geq \kappa_n$ for all $i \in [k'']$, where $\kappa_n := \frac{1}{n}\lfloor n\kappa \rfloor$.

We would like to increase p''_i to κ whenever it is smaller than that, while compensating for this by decreasing those probabilities that are larger than κ. However, there is a potential obstruction, namely that $k''\kappa$ could be greater than 1, in which case it is clearly impossible to increase all k'' probabilities to at least κ. For comparison, we know that $k''\kappa_n \leq 1$, but that is a slightly weaker assertion.

To deal with this difficulty, we will show that there exist some n_0 depending on κ such that for $n \geq n_0$, we do have $\kappa k'' \leq 1$. First, note that $\kappa_n > \kappa - \frac{1}{n}$. Thus,

$$k'' \leq \left\lfloor \frac{1}{\kappa - 1/n} \right\rfloor.$$

As $n \to \infty$, $1/(\kappa - 1/n)$ approaches $1/\kappa$ from above, and thus

$$\left\lfloor \frac{1}{\kappa - 1/n} \right\rfloor = \left\lfloor \frac{1}{\kappa} \right\rfloor$$

for all sufficiently large n. If we take n_0 to be sufficiently large, then for $n \geq n_0$ we have

$$k''\kappa \leq \left\lfloor \frac{1}{\kappa} \right\rfloor \kappa \leq 1.$$

Given this, we now define $W' = (\mathbf{p}, B)$ as follows: let I_- be the set of indices $i \in [k'']$ such that $p''_i < \kappa$, and let $\delta = \sum_{i \in I_-} (\kappa - p''_i)$. For $i \in I_-$, we then set $p_i = \kappa$, while for $i \notin I_-$ we first decrease the largest p''_i until we either hit κ or have used up the excess δ. If we stop because we hit κ, then we move to the next largest p''_i, etc. Since in the second step, we will eventually use up the excess δ, this process constructs a distribution $\mathbf{p}$ such that $p_i \geq \kappa$ for all $i \in [k'']$, and such that $\sum_i |p''_i - p_i| = 2\delta$. Note for future reference that $\delta \leq k''/n$.

Writing the embedding $\mathsf{W}[W'']$ of W'' into the set of graphons over $[0,1]$ as $\sum_{i,j} B_{ij} 1_{Y_i'' \times Y_j''}$, we construct corresponding measurable sets Y_i such that $Y_1, \dots, Y_{k''}$ forms a partition of $[0,1]$ with $\lambda(Y_i) = p_i$ and $\lambda(Y_i \triangle Y_i'') \leq |p_i - p_i''|$. (Each set Y_i will be either a superset or a subset of Y_i'', according to whether p_i'' was increased or decreased.)

For $i \in I_-$,

$$\lambda(Y_i \triangle Y_i'') \leq |p_i - p_i''| \leq \frac{1}{n} \leq \frac{1}{\kappa_n n} \lambda(Y_i'').$$

For $i \notin I_-$,

$$\lambda(Y_i \triangle Y_i'') \leq |p_i - p_i''| \leq \delta \leq \frac{k''}{n} \leq \frac{1}{\kappa^2 n} \lambda(Y_i'').$$

When n is sufficiently large, $\kappa_n \geq \kappa^2$. Increase n_0 enough for this to hold, as well as $n_0 \geq 1/\kappa^2$. Then for $n \geq n_0$,

$$\delta_p(W', W'') \leq \sqrt[p]{\frac{2^{p+1}}{\kappa^2 n}} \|W''\|_p \leq 2 \sqrt[p]{\frac{2^{p+1}}{\kappa^2 n}} \|W\|_p$$

by Lemma C.4, as in the proof of the first bound. This concludes the proof of the second bound. □

C.2. Convergence of *W*-weighted graphs. Recall that by Theorem 3.10, the sequence $G_n = G_n(\rho_n W)$ converges to W in the cut metric if $W \in L^p$ is normalized, $n\rho_n \to \infty$, and $\rho_n \to 0$. Our next lemma, which is a slight strengthening of Theorem 2.14(a) from [2], states that for the weighted graphs $Q_n(\rho_n W)$, the same holds in the tighter distance δ_p. Recalling that for any graphon, we can find an equivalent graphon over $[0,1]$, we will restrict ourselves to the case where W is a graphon over $[0,1]$, in which case we can use an even tighter distance, the distance $\hat{\delta}_p$ defined in (3.5).

LEMMA C.6. *Let $p \geq 1$, let W be a normalized L^p graphon over $[0,1]$, let $x_1, x_2, \dots \in [0,1]$ be chosen i.i.d. uniformly at random, and let ρ_n be a sequence of positive numbers such that $\rho_n \to 0$. Given $n \geq 2$, let Q_n be the $n \times n$ matrix with entries $\min\{1, \rho_n W(x_i, x_j)\}$, relabeled in such a way that $x_1 < x_2 < \cdots < x_n$. Then a.s. $\|\frac{1}{\rho_n}\mathsf{W}[Q_n] - W\|_p \to 0$, so in particular $\rho(Q_n)/\rho_n \to 1$ and $\hat{\delta}_p(\frac{1}{\rho_n} Q_n, W) \to 0$.*

PROOF. We first note that the statement clearly holds if W is replaced by the block model $W^{(k)} = W_{\mathcal{P}_k}$, where $\mathcal{P}_k$ is the partition of $[0,1]$ into consecutive intervals of length $1/k$. To see this, one just needs to use the fact that as $n \to \infty$, the fraction of points x_i which fall into the j^{th} interval converges a.s. to $1/k$.

To prove the lemma for general W, we will use Lemma C.3. Let $\rho = \rho_n$, fix $\varepsilon > 0$, choose k so that $\|W - W^{(k)}\|_p \le \varepsilon$, and let M be large enough that $\|W 1_{W \ge M}\|_p \le \varepsilon$. Also, define $W_\rho = \min\{W, 1/\rho\}$. Noting that $\frac{1}{\rho} Q_n = H_n(W_\rho)$, we then bound

$$\begin{aligned}
\|W - \frac{1}{\rho} Q_n\|_p &= \|W - H_n(W_\rho)\|_p \\
&\le \|W - W^{(k)}\|_p + \|W^{(k)} - H_n(W^{(k)})\|_p \\
&\quad + \|H_n(W^{(k)}) - H_n(W)\|_p + \|H_n(W) - H_n(W_\rho)\|_p.
\end{aligned}$$

Assuming n is large enough to ensure that $\rho^{-1} \ge M$ (which in turn implies that $|W - W_\rho| = W - W_\rho \le W 1_{W \ge M}$), we then bound the right side by

$$\varepsilon + \|W^{(k)} - H_n(W^{(k)})\|_p + \|H_n(W^{(k)} - W)\|_p + \|H_n(W 1_{W \ge M})\|_p.$$

As $n \to \infty$, the second term tends to zero with probability 1, and the third and the fourth both converge to quantities which are at most ε by Lemma C.3. Thus, with probability 1, the limit superior of $\|W - \frac{1}{\rho} Q_n\|_p$ is at most 3ε. Since ε was arbitrary, this proves the claim. □

We also formulate a quantitative version of Lemma C.6 for block models.

Lemma C.7. *Let C be a positive real number, let $\kappa \in (0, 1)$, and let W' be a block model with minimal class size at least κ, represented as a graphon over $[0, 1]$. If $\frac{1}{n\kappa} \log n \le C$, then*

$$\hat{\delta}_p(H_n(W'), W') = O_p\left(\sqrt[2p]{\frac{\log n}{n\kappa}}\right) \|W'\|_p,$$

and if $\kappa = \kappa_n$ is such that $\limsup_n \frac{1}{\kappa n} \log n < C$, then with probability 1, there exists a random n_0 such that for $n \ge n_0$,

$$\hat{\delta}_p(H_n(W'), W') = O\left(\sqrt[2p]{\frac{\log n}{n\kappa}}\right) \|W'\|_p.$$

Here the constants implicit in the big-O and O_p symbols depend only on C.

To prove Lemma C.7 we prove the following adaptation of a lemma from [5], where only bounded graphons where considered.

LEMMA C.8. *Let $\mathcal{P} = (Y_1, \dots, Y_k)$ be a partition of $[0,1]$ into consecutive intervals, and let W be a graphon over $[0,1]$ that is constant on sets of the form $Y_i \times Y_j$. If $x_1, \dots, x_n$ are chosen i.i.d. uniformly at random from $[0,1]$ and H_n is the $n \times n$ matrix with entries $W(x_i, x_j)$, then*

$$\hat{\delta}_p(H_n, W) \le 2\sqrt[p]{\varepsilon(1+\varepsilon)}\|W\|_p,$$

where ε is the random variable

$$\varepsilon = \max_{i \in [k]} \frac{1}{\lambda(Y_i)}\Big(\frac{1}{n} + \Big|\frac{n_i}{n} - \lambda(Y_i)\Big|\Big),$$

with n_i denoting the number of points x_ℓ that lie in Y_i.

PROOF. Let $I_1, \dots, I_n$ be a partition of $[0,1]$ into adjacent intervals of lengths $1/n$. Then $\mathsf{W}[H_n]$ is of the form $\sum_{i,j} B_{ij} 1_{Y_i' \times Y_j'}$, where Y_i' is the union of n_i of the intervals $I_1, \dots, I_n$ (which particular n_i intervals depends on the labeling of the vertices of H_n). In fact, given a map $\tau\colon [n] \to [k]$, define $Y_i' = Y_i'(\tau)$ to be the union of all intervals I_ℓ such that $\tau(\ell) = i$, and let $W(\tau) = \sum_{i,j} B_{ij} 1_{Y_i'(\tau) \times Y_j'(\tau)}$. Then

$$\hat{\delta}_2(H_n, W) = \min_{\tau} \|W(\tau) - W\|_2,$$

where the minimum is over all τ such that $|\tau^{-1}(\{i\})| = n_i$ for all i. In view of Lemma C.4, we will want to keep the Lebesgue measure of $Y_i \triangle Y_i'$ small for all i. We claim that this is indeed possible, and that τ can be chosen in such a way that

$$\lambda(Y_i \triangle Y_i') \le \Big|\frac{n_i}{n} - \lambda(Y_i)\Big| + \frac{1}{n} \quad \text{for all } i. \tag{C.4}$$

To prove this, we note that choosing τ is equivalent to choosing, for all i, n_i of the intervals $I_1, \dots, I_n$ to make up Y_i'.

Let $\tilde{Y}_1, \dots, \tilde{Y}_k$ be obtained from $Y_1, \dots, Y_k$ by rounding the endpoints to the nearest integer multiples of $1/n$, choosing the multiple to the left in case of a tie. With this convention,

$$\lfloor \lambda(Y_i)n \rfloor \le \lambda(\tilde{Y}_i)n \le \lceil \lambda(Y_i)n \rceil.$$

Thus, if $n_i \le \lambda(Y_i)n$, then $n_i \le n\lambda(\tilde{Y}_i)$, while if $n_i \ge \lambda(Y_i)n$, then $n_i \ge n\lambda(\tilde{Y}_i)$. Keeping this in mind, we see that for $n_i \le \lambda(Y_i)n$, we can find at least n_i intervals I_ℓ that, except possibly for their endpoints, are subsets of $\tilde{Y}_i$. We will define Y_i' to be the union of these intervals. In a similar way, if $n_i > \lambda(Y_i)n$, we choose $n\lambda(\tilde{Y}_i) \leq n_i$ intervals (namely, those forming $\tilde{Y}_i$) to build a preliminary set $Y_i^{(0)}$. Having done this for all i, we take a second run through all i with $n_i > \lambda(Y_i)n$, choosing an arbitrary set of $n_i - \lambda(\tilde{Y}_i)n$ intervals I_ℓ from those not yet assigned at this point. At the end of this round, we end up with sets Y_i' such that Y_i' is the union of n_i intervals from $I_1, \dots, I_n$, with the additional property that

$$\text{either} \quad Y_i' \subseteq \tilde{Y}_i \quad \text{or} \quad \tilde{Y}_i \subseteq Y_i'.$$

But this implies that $\lambda(Y_i' \triangle \tilde{Y}_i) = |\frac{n_i}{n} - \lambda(\tilde{Y}_i)|$ for all i. Since the endpoints of Y_i get shifted by at most $1/(2n)$ in order to obtain $\tilde{Y}_i$, the additional error in going from $\tilde{Y}_i$ to Y_i is at most $1/n$, proving (C.4). Combined with Lemma C.4, this concludes the proof. $\square$

Finally, the following lemma implies Lemma C.7.

LEMMA C.9. *Let ε and the other notation be as in Lemma C.8, suppose that all sizes of $\mathcal{P}$ have measure at least κ, and let $\eta \in (0,1)$. Then*

$$\varepsilon \leq \frac{1}{\kappa n} + \max\left\{\frac{3}{n\kappa}\log\frac{2}{\kappa\eta}, \sqrt{\frac{3}{n\kappa}\log\frac{2}{\kappa\eta}}\right\}$$

with probability at least $1-\eta$. As a consequence, if C is a positive real number, then

$$\hat{\delta}_p(H_n, W) = O_p\left(\sqrt[2p]{\frac{\log n}{n\kappa}}\right)\|W\|_p$$

whenever $\frac{\log n}{n\kappa} \leq C$, with the constant implicit in the O_p symbol depending on C. In addition, if $\kappa = \kappa_n$ is such that $\limsup_n \frac{1}{\kappa n}\log n < C$, then with probability 1, there exists a random n_0 such that for $n \geq n_0$,

$$\hat{\delta}_p(H_n, W) = O\left(\sqrt[2p]{\frac{\log n}{n\kappa}}\right)\|W\|_p,$$

with the constant implicit in the big-O symbol again depending on C.

PROOF. By the multiplicative Chernoff bound,

$$\begin{aligned}\Pr\left(\left|\frac{n_i}{n} - \lambda(Y_i)\right| \geq t\lambda(Y_i)\right) &\leq 2\exp\left(-\frac{n\lambda(Y_i)}{3}\min\{t,t^2\}\right)\\ &\leq 2\exp\left(-\frac{n\kappa}{3}\min\{t,t^2\}\right),\end{aligned}$$

so by the union bound and the fact that the number k of classes is at most $1/\kappa$, we get

$$\varepsilon \le t + \frac{1}{\kappa n} \quad \text{with probability at least} \quad 1 - \frac{2}{\kappa}\exp\left(-\frac{n\kappa}{3}\min\{t, t^2\}\right).$$

Setting $y = \frac{3}{n\kappa}\log\frac{2}{\kappa\eta}$ we see that with probability at least $1-\eta$, $\varepsilon \le t + \frac{1}{\kappa n}$ whenever $\min\{t, t^2\} \ge y$. This implies the bound on ε.

For the remaining part of the proof, choose $\eta = 2n^{-2}$. Then with probability at least $1 - 2n^{-2}$,

$$\begin{aligned}
\varepsilon &\le \frac{1}{\kappa n} + \max\left\{\frac{3}{n\kappa}\log\frac{n^2}{\kappa}, \sqrt{\frac{3}{n\kappa}\log\frac{n^2}{\kappa}}\right\} \\
&\le \frac{1}{\kappa n} + \max\left\{\frac{9}{n\kappa}\log 2Cn, \sqrt{\frac{9}{n\kappa}\log 2Cn}\right\} \quad \left(\text{because } \frac{1}{n\kappa} \le \frac{C}{\log n} \le 2C\right) \\
&\le \sqrt{\frac{C'' \log n}{n\kappa}} \le \sqrt{CC''},
\end{aligned}$$

for some C'' depending on C. This implies

$$2^p\varepsilon(1+\varepsilon) \le 2^p(1+\sqrt{CC''})\sqrt{\frac{C''\log n}{n\kappa}} =: C'\sqrt{\frac{\log n}{n\kappa}}$$

and hence

$$\hat{\delta}_p(H_n, W) \le \sqrt[2p]{\frac{C'\log n}{n\kappa}}\|W\|_p.$$

Since the failure probability $2n^{-2}$ is summable, this implies the a.s. statement. To prove the statement in probability, we note that by Remark C.5, $\hat{\delta}_p(H_n, W) \le 2\kappa^{-2/p}\|W\|_p$ always. If we combine this inequality with the previous bound (which holds with probability $1-\eta$), we find that

$$\begin{aligned}
\mathbb{E}\left[(\hat{\delta}_p(H_n, W))^p\right] &\le \left(\sqrt{\frac{C'\log n}{n\kappa}}(1-\eta) + \frac{2^p}{\kappa^2}\eta\right)\|W\|_p^p \\
&= \left(\sqrt{\frac{C'\log n}{n\kappa}} + \frac{2^{p+1}}{\kappa^2 n^2}\right)\|W\|_p^p \\
&= O\left(\sqrt{\frac{\log n}{n\kappa}}\right)\|W\|_p^p.
\end{aligned}$$

This implies the O_p bound for $\hat{\delta}_p(H_n, W)$. $\square$

C.3. Proof of Lemma 6.2.

Proof of Lemma 6.2. Without loss of generality we can assume that W is continuous, because continuous functions are dense in L^1 and $||W_{\mathcal{P}_n} - W'_{\mathcal{P}_n}||_1 \le ||W - W'||_1$.

Let $J_1, \dots, J_N$ be the parts of $\mathcal{P}_n$. Then for $(x, y) \in J_i \times J_j$,

$$W_{\mathcal{P}_n}(x, y) = \frac{1}{\lambda(J_i)\lambda(J_j)} \int_{J_i \times J_j} W(u, v)\, du\, dv.$$

By combining this formula with

$$W(x, y) = \frac{1}{\lambda(J_i)\lambda(J_j)} \int_{J_i \times J_j} W(x, y)\, du\, dv,$$

we find that

$$\|W_{\mathcal{P}_n} - W\|_1 \le \sum_{i,j=1}^{N} \frac{1}{\lambda(J_i)\lambda(J_j)} \int_{J_i \times J_j} \int_{J_i \times J_j} |W(u, v) - W(x, y)|\, du\, dv\, dx\, dy.$$

Because W is continuous on $[0, 1]^2$ (and hence uniformly continuous), for each $\delta > 0$, there exists $\varepsilon > 0$ such that $|W(x, y) - W(u, v)| < \delta$ whenever $|x - u| < \varepsilon$ and $|y - v| < \varepsilon$. Then

$$\begin{aligned}
&||W_{\mathcal{P}_n} - W||_1 \\
&\le \delta + \sum_{i,j=1}^{N} \frac{2\|W\|_\infty}{\lambda(J_i)\lambda(J_j)} \int_{J_i \times J_j} \int_{J_i \times J_j} 1_{|x-u| \ge \varepsilon \text{ or } |y-v| \ge \varepsilon}\, du\, dv\, dx\, dy \\
&\le \delta + \sum_{i,j=1}^{N} \frac{4\|W\|_\infty}{\lambda(J_i)\lambda(J_j)} \int_{J_i \times J_j} \int_{J_i \times J_j} 1_{|x-u| \ge \varepsilon}\, du\, dv\, dx\, dy \\
&= \delta + 4\|W\|_\infty \sum_{i=1}^{N} \frac{1}{\lambda(J_i)} \int_{J_i \times J_i} 1_{|x-u| \ge \varepsilon}\, du\, dx \\
&= \delta + 4\|W\|_\infty p_{n,\varepsilon}.
\end{aligned}$$

It follows that

$$\limsup_{n \to \infty} ||W_{\mathcal{P}_n} - W||_1 \le \delta$$

for each $\delta > 0$, as desired. □

APPENDIX D: PROOF OF THEOREM 2.1

Proof of Theorem 2.1. Let $(\Omega, \mathcal{F}, \pi)$ be the space on which W is defined, and let $Q = Q_n(\rho W)$ as before. Defining $W_\rho = \min\{W, 1/\rho\}$, we will write Q as $\rho H_n(W_\rho)$ and $\mathrm{tail}_\rho^{(2)}(W) = \|W - W_\rho\|_2$.

By the triangle inequality and the fact that the $\hat{\delta}_2$ distance dominates the δ_2 distance, we have

$$\begin{aligned} \delta_2\left(\frac{1}{\rho}\widehat{W}, W\right) &= \delta_2\left(M_n\left(\frac{1}{\rho}\widehat{W}\right), W\right) \\ &\le \hat{\delta}_2\left(M_n\left(\frac{1}{\rho}\widehat{W}\right), \frac{1}{\rho}Q\right) + \delta_2\left(\frac{1}{\rho}Q, W\right). \end{aligned} \tag{D.1}$$

To bound the first term on the right side, we will use Theorem 4.1 and then bound $\hat{\varepsilon}_{\geq\kappa}^{(2)}\left(\frac{1}{\rho}Q\right)$ in terms of $\varepsilon_{\geq\kappa}^{(2)}(W)$.

Recall that by Lemma C.1 the infimum in the definition (2.1) of $\varepsilon_{\geq\kappa}^{(2)}(W)$ is a minimum, and the minimizer $W' \in \mathcal{B}_{\geq\kappa}$ satisfies $\|W'\|_2 \le 2\|W\|_2$. As established in Lemma C.2, we can relabel the blocks of the block model W' in such a way that

$$\hat{\delta}_2(W'', \mathsf{W}[W']) \le \sqrt{\frac{8}{\kappa n}}\|W'\|_2 \le 2\sqrt{\frac{8}{\kappa n}}\|W\|_2 = \sqrt{\frac{32}{\kappa n}}\|W\|_2$$

for some $W'' \in \mathcal{A}_{n,\geq\kappa}$. Setting $\widetilde{W}' = \mathsf{W}[W']$, we find that

$$\begin{aligned} \hat{\varepsilon}_{\geq\kappa}^{(2)}\left(\frac{1}{\rho}Q\right) &\le \hat{\delta}_2\left(\frac{1}{\rho}Q, W''\right) \\ &\le \hat{\delta}_2\left(\frac{1}{\rho}Q, \widetilde{W}'\right) + \sqrt{\frac{32}{\kappa n}}\|W\|_2 \\ &= \hat{\delta}_2\left(H_n(W_\rho), \widetilde{W}'\right) + \sqrt{\frac{32}{\kappa n}}\|W\|_2. \end{aligned}$$

Next we would like to choose a coupling μ of $\mathbf{p}$ and π such that

$$\varepsilon_{\geq\kappa}^{(2)}(W) = \delta_2(W', W) = \|W' - W\|_{2,\mu},$$

where $\|\cdot\|_{2,\mu}$ denotes the L^2 norm with respect to the coupling μ. (This is an abuse of notation, but it is more convenient than writing out the formula, as in (1.1).) Such a coupling need not exist, but that is not a significant obstacle. We could complete the proof by looking at couplings that come arbitrarily close to the oracle error, but instead we will switch to equivalent graphons over $[0,1]$, because Lemma A.7 then guarantees the existence of an optimal coupling. The oracle error and tail bounds are invariant under equivalence, so we can assume without loss of generality that the coupling μ exists.

We use this coupling to couple the random graphs $Q(\rho W)$ and $Q(\rho W')$. With the help of the triangle inequality, we then conclude that

$$\begin{aligned}\hat{\varepsilon}^{(2)}_{\geq\kappa}\left(\frac{1}{\rho}Q\right) &\leq \|H_n(W_\rho)-H_n(W)\|_2+\|H_n(W)-H_n(W')\|_2\\ &\quad+\hat{\delta}_2\Big(H_n(W'),\widetilde{W'}\Big)+\sqrt{\frac{32}{\kappa n}}\|W\|_2.\end{aligned}\tag{D.2}$$

After these preparations, we start with the proof of (i). To this end, we first use the triangle inequality and the fact that $\delta_2(W',W)=\varepsilon^{(2)}_{\geq\kappa}(W)$ to bound

$$\begin{aligned}\delta_2\left(\frac{1}{\rho}Q,W\right) &= \delta_2(H_n(W_\rho),W)\\ &\leq \|H_n(W_\rho)-H_n(W)\|_2+\|H_n(W)-H_n(W')\|_2\\ &\quad+\delta_2(H_n(W'),W')+\varepsilon^{(2)}_{\geq\kappa}(W).\end{aligned}$$

Next we estimate

$$\begin{aligned}\mathbb{E}\left[\|H_n(W_\rho)-H_n(W)\|_2\right] &= \mathbb{E}\left[\|H_n(W_\rho-W)\|_2\right] \leq \sqrt{\mathbb{E}\left[\|H_n(W_\rho-W)\|_2^2\right]}\\ &= \|W_\rho-W\|_2=\mathrm{tail}^{(2)}_\rho(W)\end{aligned}$$

and

$$\mathbb{E}\left[\|H_n(W)-H_n(W')\|_2\right]\leq\|W-W'\|_{2,\mu}=\varepsilon^{(2)}_{\geq\kappa}(W).$$

Since $\hat{\delta}_2\Big(H_n(W'),\widetilde{W'}\Big)$ has the same distribution as $\hat{\delta}_2\Big(H_n(\widetilde{W'}),\widetilde{W'}\Big)$, we may then use Lemma C.7 and the fact that $\|W'\|_2\leq 2\|W\|_2$ to conclude that

$$\hat{\varepsilon}^{(2)}_{\geq\kappa}\left(\frac{1}{\rho}Q\right)=O_p\left(\mathrm{tail}^{(2)}_\rho(W)+\varepsilon^{(2)}_{\geq\kappa}(W)+\sqrt[4]{\frac{\log n}{n\kappa}}\|W\|_2\right).$$

(Note that $(1+\log(1/\kappa))\kappa^{-2}=O(\rho n)$ implies that $1/\sqrt{n}=O(\kappa)$ and hence $\log n=O(\kappa n)$, as required for the application of Lemma C.7.) In a similar way, we use the fact that $\delta_2(H_n(W'),W')$ has the same distribution as $\delta_2(H_n(\widetilde{W'}),\widetilde{W'})$ to conclude that

$$\delta_2\left(\frac{1}{\rho}Q,W\right)=O_p\left(\mathrm{tail}^{(2)}_\rho(W)+\varepsilon^{(2)}_{\geq\kappa}(W)+\sqrt[4]{\frac{\log n}{n\kappa}}\|W\|_2\right).$$

With the help of (D.1) and Theorem 4.1, this implies that

$$\delta_2\left(\frac{1}{\rho}\widehat{W}, W\right) = O_p\left(\mathrm{tail}_\rho^{(2)}(W) + \varepsilon_{\geq\kappa}^{(2)}(W) + \sqrt[4]{\frac{\log n}{n\kappa}}\|W\|_2 \right.$$
$$\left. + \sqrt[4]{\frac{1+\log(1/\kappa)}{\kappa^2\rho n}}\right),$$

which concludes the proof of (i).

Next we prove (ii). Since W is square integrable, $\|W - W_\rho\|_2 \to 0$ as $\rho \to 0$, so by combining the law of large numbers for U-statistics (see Lemma C.3 in Appendix C) with a simple two ε argument, we conclude that a.s., the first term in (D.2) tends to zero. Again by the law of large numbers for U-statistics, the second term tends to $\|W' - W\|_{2,\mu} = \varepsilon_{\geq\kappa}^{(2)}(W)$, and by Lemma C.7 and the fact that $H_n(W')$ and $H_n(\widetilde{W'})$ have the same distribution, the third term tends to zero as well. Thus a.s., the right side of (D.2) tends to $\varepsilon_{\geq\kappa}^{(2)}(W)$. Combined with (D.1), Lemma C.6, and Theorem 4.1, we see that for fixed κ,

$$\limsup_{n\to\infty} \delta_2\left(\frac{1}{\rho}\widehat{W}, W\right) \leq \varepsilon_{\geq\kappa}^{(2)}(W) \quad \text{with probability 1.}$$

On the other hand, by the second bound in Lemma C.2,

$$\varepsilon_{\geq\kappa}^{(p)}(W) \leq \liminf_{n\to\infty} \min_{W''\in\mathcal{B}_{\geq\kappa,n}} \delta_p(W'', W).$$

Since $\frac{1}{\rho}\widehat{W} \in \mathcal{B}_{\geq\kappa,n}$, this yields $\varepsilon_{\geq\kappa}^{(2)}(W) \leq \liminf_{n\to\infty} \delta_2\left(\frac{1}{\rho}\widehat{W}, W\right)$, completing the proof of (ii).

To prove (iii), note that the condition $\kappa_n^{-2}\log(1/\kappa_n) = o(n\rho_n)$ implies in particular that $\kappa_n\sqrt{n} \to \infty$, which in turn implies that $\frac{1}{\kappa_n n}\log n \to 0$. We may therefore again use Lemma C.7 to show that the third term in (D.2) tends to zero a.s. The first term does not depend on κ, and hence tends to zero just as before, but now the second term tends to zero as well, by a two ε argument invoking now the fact that $\|W' - W\|_{2,\mu} = \varepsilon_{\geq\kappa_n}^{(2)}(W) \to 0$. Since the condition $\kappa_n^{-2}\log(1/\kappa_n) = o(n\rho_n)$ clearly implies that $n\kappa_n \to \infty$, we conclude that a.s., $\hat{\varepsilon}_{\geq\kappa_n}^{(2)}\left(\frac{1}{\rho}Q\right) \to 0$. Combined with (D.1), Lemma C.6, and Theorem 4.1, this implies (iii). $\square$

APPENDIX E: PROOFS OF THEOREMS 5.1 AND 2.2

PROOF OF THEOREM 5.1. Let $A = A(G)$ and $k = \lceil \frac{n}{\lfloor \kappa n \rfloor} \rceil$. We will show that

$$\hat{\delta}_\square\Big(M_n(\widehat{W}), Q\Big) \le 2\hat{\varepsilon}_{\ge\kappa,\square}(Q) + 3\|Q - A\|_\square. \tag{E.1}$$

To this end, we first prove that

$$\|M_n(\widehat{W}) - A\|_\square \le 2 \min_{M \in \mathcal{A}_{n,\ge\kappa}} \|M - A\|_\square. \tag{E.2}$$

To see this, we note that $\mathcal{A}_{n,\ge\kappa}$ consists of all $n \times n$ matrices M such that $M = M_\tau$ for some $\tau\colon [n] \to [k]$ such that the smallest non-empty class of τ has at least size $\lfloor \kappa n \rfloor$. Next we observe that for all $\tau\colon [n] \to [k]$, the map $H \mapsto H_\tau$ is a contraction in the cut norm. As a consequence, for all $n \times n$ matrices M with $M = M_\tau$,

$$\|A_\tau - A\|_\square \le \|A_\tau - M_\tau\|_\square + \|M - A\|_\square \le 2\|M - A\|_\square.$$

Because $M_n(\widehat{W}) = A_{\hat\tau}$ for some $\hat\tau\colon [n] \to [k]$ that minimizes $\|A - A_\tau\|_\square$ over all τ whose smallest non-empty class has size at least $\lfloor \kappa n \rfloor$, the bound (E.2) now follows.

After this preparation, the proof of (E.1) is straightforward. Indeed,

$$\begin{aligned}
\hat{\delta}_\square\Big(M_n(\widehat{W}), Q\Big) &\le \|M_n(\widehat{W}) - A\|_\square + \|A - Q\|_\square \\
&\le 2 \min_{M \in \mathcal{A}_{n,\ge\kappa}} \|M - A\|_\square + \|A - Q\|_\square \\
&\le 2 \min_{M \in \mathcal{A}_{n,\ge\kappa}} \|M - Q\|_\square + 3\|A - Q\|_\square \\
&= 2\hat{\varepsilon}_{\ge\kappa,\square}(Q) + 3\|Q - A\|_\square.
\end{aligned}$$

From here on, the proof proceeds along the same lines as that of Theorem 4.1, this time starting from the observation (5.1). Using this fact and a concentration argument, we now can show that conditioned on Q, if $\rho(Q)n \ge 1$ then

$$\|Q - A\|_\square \le 15\sqrt{\frac{\rho(Q)}{n}}$$

holds with probability at least $1 - e^{-n}$, and

$$\|Q - A\|_\square = O_p\left(\sqrt{\frac{\rho(Q)}{n}}\right),$$

independently of the condition $\rho(Q)n \ge 1$; see Lemma H.3 in Appendix H for details. The assertions of the theorem now follow. $\square$

Proof of Theorem 2.2. Keeping the notation from the proof of Theorem 2.1, and using the fact that the distance $\hat{\delta}_\square$ is dominated by the distance $\hat{\delta}_1$, we now bound

$$\delta_\square\left(\frac{1}{\rho}\widehat{W}, W\right) \le \hat{\delta}_\square\left(M_n\left(\frac{1}{\rho}\widehat{W}\right), \frac{1}{\rho}Q\right) + \delta_1\left(\frac{1}{\rho}Q, W\right). \tag{E.3}$$

By Lemma C.1 and Lemma C.2 for $p = 1$,

$$\hat{\varepsilon}_{\ge\kappa,\square}\left(\frac{1}{\rho}Q\right) \le \hat{\delta}_1\left(\frac{1}{\rho}Q, \widetilde{W}'\right) + \frac{8}{\kappa n} = \hat{\delta}_1\left(H_n(W_\rho), \widetilde{W}'\right) + \frac{8}{\kappa n},$$

where W' is a minimizer for (2.1) for $p = 1$, with $\|W'\|_1 \le 2\|W\|_1 = 2$, and $\widetilde{W}'$ again stands for $\mathsf{W}[W']$. Writing $\varepsilon^{(1)}_{\ge\kappa}(W)$ as $\varepsilon^{(1)}_{\ge\kappa}(W) = \delta_1(W', W) = \|W' - W\|_{1,\mu}$ for some coupling μ of $\mathbf{p}$ and π (which we can assume exists without loss of generality by passing to equivalent graphons over $[0, 1]$, as in the proof of Theorem 2.1), we then get

$$\begin{aligned}\hat{\varepsilon}_{\ge\kappa,\square}\left(\frac{1}{\rho}Q\right) \le{}& \|H_n(W_\rho) - H_n(W)\|_1 + \|H_n(W) - H_n(W')\|_1 \\ &+ \hat{\delta}_1\left(H_n(W'), \widetilde{W}'\right) + \frac{8}{\kappa n}\end{aligned}$$

and

$$\begin{aligned}\delta_1\left(\frac{1}{\rho}Q, W\right) \le{}& \|H_n(W_\rho) - H_n(W)\|_1 + \|H_n(W) - H_n(W')\|_1 \\ &+ \delta_1(H_n(W'), W') + \varepsilon^{(1)}_{\ge\kappa}(W),\end{aligned}$$

where as before $H_n(W)$ and $H_n(W')$ are coupled with the help of μ. From here on, the proof of Theorem 2.2 proceeds exactly as the proof of Theorem 2.1 did, with the condition $\frac{1}{n\kappa}\log n = O(1)$ that is needed to apply Lemma C.7 guaranteed by the hypotheses of the theorem. We finally arrive at

$$\hat{\varepsilon}_{\ge\kappa,\square}\left(\frac{1}{\rho}Q\right) = O_p\left(\mathrm{tail}^{(1)}_\rho(W) + \varepsilon^{(1)}_{\ge\kappa}(W) + \sqrt{\frac{\log n}{n\kappa}}\right)$$

and

$$\delta_1\left(\frac{1}{\rho}Q, W\right) = O_p\left(\mathrm{tail}^{(1)}_\rho(W) + \varepsilon^{(1)}_{\ge\kappa}(W) + \sqrt{\frac{\log n}{n\kappa}}\right).$$

With the help of (E.3) and Theorem 5.1, this proves the bound in probability.

The almost sure statements are proved similarly. $\square$

APPENDIX F: PROOFS OF PROPOSITIONS 7.1 AND 7.2

The proof of Proposition 7.1 relies on the following technical lemma.

LEMMA F.1. *Let W be a bounded graphon over some probability space $(\Omega, \mathcal{F}, \pi)$, and let W' be a graphon over $(\Omega, \mathcal{F}, \pi)$ such that $\|W - W'\|_p \le \varepsilon$ and W' is a block model with N classes. Then*

$$\varepsilon^{(p)}_{\ge\kappa}(W) \le 2\varepsilon \qquad \textit{whenever} \qquad \kappa \le \frac{1}{2N}\Big(\frac{\varepsilon}{\|W'\|_\infty}\Big)^p.$$

PROOF. Suppose W' is based on the partition $(Y_1, \dots, Y_N)$ of Ω. Arranging the classes Y_i in $\mathcal{P}$ in order of decreasing measure, let Y_ℓ be the last class of measure κ or more. We then define $Y'_\ell = \bigcup_{i\ge\ell} Y_i$, and $Y'_i = Y_i$ for all $i < \ell$. Let W'' be a block model with blocks $Y'_1, \dots, Y'_\ell$ and the same values as W' on $Y_i \times Y_j$ when $i, j < \ell$ but the value 0 when i or j equals ℓ. Clearly $W'' \in \mathcal{B}_{\ge\kappa}$. To prove the proposition, we will have to show that $\|W' - W''\|_p \le \varepsilon$. To this end, we note that W'' and W' agree on $\Omega_0 \times \Omega_0$, where $\Omega_0 = Y_1 \cup \dots \cup Y_{\ell-1}$, and that $\|W' - W''\|_\infty \le \|W'\|_\infty$. As a consequence,

$$\|W' - W''\|_p = \|(W' - W'')(1 - 1_{\Omega_0\times\Omega_0})\|_p \le \|W'\|_\infty\big(1 - \pi(\Omega_0)^2\big)^{1/p}.$$

But because the classes $Y_{\ell+1}, \dots, Y_N$ have measure smaller than κ,

$$\pi(\Omega_0) \ge 1 - \ell\kappa \ge 1 - N\kappa,$$

showing that

$$\|W' - W''\|_p \le \|W'\|_\infty\big(2N\kappa\big)^{1/p},$$

which is bounded by ε if $\kappa \le \frac{1}{2N}(\varepsilon/\|W'\|_\infty)^p$. □

PROOF OF PROPOSITION 7.1. We will prove the proposition for $D = 1 + 2C(2R)^\alpha$.

To prove the first statement, let $C_0 = \min_{x,y\in\Lambda_R} W(x, y)$. Since $C_0 = \int C_0 \le \|W\|_1 = 1$, Hölder continuity implies that $\|W\|_\infty \le 1 + 2C(2R)^\alpha = D$.

To prove the second statement, consider $k \in \mathbb{N}$, and let $\mathcal{P}$ be the partition of Λ_R into k^d cubes of side-length $a = 2R/k$. For a given class $Y \in \mathcal{P}$, two points $x, x' \in Y$ have distance $|x-x'|_\infty \le a$. Thus, if Y and Y' are two classes in $\mathcal{P}$, then $|W(x, y) - W(x', y')| \le 2Ca^\alpha = 2C(2Rk^{-1})^\alpha$ whenever $x, y \in Y$ and $x', y' \in Y'$. As a consequence $\|W - W_{\mathcal{P}}\|_\infty \le 2C(2Rk^{-1})^\alpha \le Dk^{-\alpha}$.

If π is the uniform measure over Λ_R, then each class Y of $\mathcal{P}$ has measure $\pi(Y) = k^{-d}$, so setting $k = \lfloor \kappa^{-1/d} \rfloor$, we obtain $k^{-1} \leq 2\kappa^{1/d}$ and thus

$$\varepsilon^{(p)}_{\geq\kappa}(W) \leq 2D\kappa^{\alpha/d},$$

which proves the proposition for the case of the uniform measure. (Recall that δ_p and hence $\varepsilon^{(p)}_{\geq\kappa}(W)$ are decreasing functions of p.)

But for general measures, some of the classes of $\mathcal{P}$ might have tiny measure. To fix this, we merge all classes of measure less than κ (where κ will now be smaller than k^{-d}) with the smallest of those which have measure at least κ. Lemma F.1 shows that for κ small enough, this actually works. To apply the lemma, we set $N = k^d$ and observe that $\|W_{\mathcal{P}}\|_\infty \leq \|W\|_\infty \leq D$ and $\|W - W_{\mathcal{P}}\|_\infty \leq DN^{-\alpha/d}$. Lemma F.1 then implies that

$$\varepsilon^{(p)}_{\geq\kappa}(W) \leq 2DN^{-\alpha/d}$$

provided $2\kappa \leq N^{-\frac{p\alpha+d}{d}}$. Thus for $\kappa \leq 1/2$, we may choose

$$k = \lfloor (2\kappa)^{-1/(p\alpha+d)} \rfloor$$

to show that (7.1) holds for $\kappa \leq 1/2$. For $\kappa \geq 1/2$, that would amount to $k = 0$, but fortunately this case is trivial: the right side of (7.1) is at least $2D$ and hence at least 2, while $\varepsilon^{(p)}_{\geq\kappa}(W) \leq 1$ for a normalized graphon, showing that (7.1) holds for $\kappa \geq 1/2$ as well. □

Proof of Proposition 7.2. Let $R_0 \geq 1$ be such that $\pi(\Lambda_{R_0}) \geq 1/2$, and let $D_0 = 4 + 2CR_0^\alpha$. Then

$$\min_{x,y\in\Lambda_{R_0}} W(x,y) \leq \frac{1}{\pi(\Lambda_{R_0})^2}\int_{\Lambda_{R_0}\times\Lambda_{R_0}} W \leq \frac{\|W\|_1}{\pi(\Lambda_{R_0})^2} \leq 4.$$

Denoting the minimizer of W in $\Lambda_{R_0} \times \Lambda_{R_0}$ by (x_0, y_0), we then have

$$W(0,0) \leq 4 + C|x_0|_\infty^\alpha + C|y_0|_\infty^\alpha \leq 4 + 2CR_0^\alpha,$$

implying that

$$W(x,y) \leq D_0 + C|x|_\infty^\alpha + C|y|_\infty^\alpha$$

for all $x, y \in \mathbb{R}^d$. It will be convenient to introduce the functions $f(x,y) = C|x|_\infty^\alpha$ and $g(x,y) = C|y|_\infty^\alpha$ and write this inequality as

$$W \leq D_0 + f + g.$$

By our definition of β' and our assumption on π,

$$\|f\|_{p(1+\beta')} = C\Big(\int |x|_\infty^\beta \, d\pi(x)\Big)^{\frac{1}{p(1+\beta')}} < \infty.$$

To prove the bound on $\operatorname{tail}_\rho^{(p)}(W)$, we observe that $0 \le W - W_\rho \le W1_{W\ge 1/\rho}$. As a consequence,

$$\begin{aligned}\operatorname{tail}_\rho^{(p)}(W) &\le \|W1_{W\ge 1/\rho}\|_p \le \rho^{\beta'}\|W^{1+\beta'}\|_p = \rho^{\beta'}\|W\|_{p(1+\beta')}^{1+\beta'}\\ &\le \rho^{\beta'}\Big(\|D_0 + f + g\|_{p(1+\beta')}\Big)^{1+\beta'}\\ &\le \rho^{\beta'}\Big(D_0 + \|f\|_{p(1+\beta')} + \|g\|_{p(1+\beta')}\Big)^{1+\beta'} \le D\rho^{\beta'}\end{aligned}$$

for some constant D depending on α, β, p, and C, as well as the measure π (via R_0 and the norm $\|f\|_{p(1+\beta')}$).

To prove the bound on the oracle error, we want to construct a good block model approximation to W. To this end, we first bound the contributions to $\|W\|_p$ that come from points x, y outside a box Λ_R, where $R \ge 1$ will be chosen later. If we set $r = CR^\alpha$, then the condition $(x, y) \notin \Lambda_R \times \Lambda_R$ implies $|x|_\infty > R$ or $|y|_\infty > R$ and hence $f + g > r$. But

$$\|W1_{f+g>r}\|_p \le r^{-\beta'}\|(f+g)^{\beta'}W\|_p \le r^{-\beta'}\|(D_0 + f + g)^{\beta'+1}\|_p \le Dr^{-\beta'},$$

and hence

$$\|W1_{(\mathbb{R}^d\times\mathbb{R}^d)\setminus(\Lambda_R\times\Lambda_R)}\|_p \le D_1 R^{-\beta'\alpha}, \tag{F.1}$$

as long as D_1 is chosen so that $D_1 \ge DC^{-\beta'}$.

Next we consider a partition $\mathcal{P} = (Y_1, \dots, Y_N)$ of Λ_R into cubes of side length $2R/k$, with $N = k^d$. We define β_{ij} to be the average of W over $Y_i \times Y_j$, and

$$W' = \sum_{i,j=1}^N \beta_{ij} 1_{Y_i\times Y_j}.$$

Since W' is composed of parts obtained by averaging over subsets in Λ_R, where W is bounded by $D_0 + 2CR^\alpha \le D_0(1 + R^\alpha) \le 2D_0R^\alpha$, we have

$$\|W'\|_\infty \le 2D_1R^\alpha,$$

provided D_1 is chosen to be at least D_0.

Inside $\Lambda_R \times \Lambda_R$, we bound $|W(x, y) - W'(x, y)|$ by

$$2C(2R/k)^\alpha = 2C(2R)^\alpha N^{-\alpha/d} \le D_1R^\alpha N^{-\alpha/d},$$

where $D_1 = \max\{D_0, DC^{-\beta'}, 2C2^\alpha\}$. Finally $W - W' = W$ outside Λ_R. Combined with the bound (F.1), we conclude that

$$\|W - W'\|_p \leq \varepsilon,$$

where $\varepsilon = D_1(R^\alpha N^{-\alpha/d} + R^{-\beta'\alpha})$. With the help of Lemma F.1 we conclude that

$$\varepsilon^{(p)}_{\geq\kappa}(W) \leq 2D_1(R^\alpha N^{-\alpha/d} + R^{-\beta'\alpha}),$$

provided that

$$\kappa \leq \frac{1}{2N}\Big(\frac{R^\alpha N^{-\alpha/d} + R^{-\beta'\alpha}}{2R^\alpha}\Big)^p = \frac{1}{2N}\Big(\frac{N^{-\alpha/d} + R^{-(\beta'+1)\alpha}}{2}\Big)^p$$

and $R \geq 1$. Choosing $R = N^{\frac{1}{d(\beta'+1)}}$, we find that

$$\varepsilon^{(p)}_{\geq\kappa}(W) \leq 4D_1 N^{-\frac{\beta'\alpha}{d(1+\beta')}},$$

provided that $\kappa \leq \frac{1}{2}N^{-\frac{p\alpha+d}{d}}$.

Because $\kappa \leq 1/2$, we can choose $k = \left\lfloor \left(\frac{1}{2\kappa}\right)^{\frac{1}{p\alpha+d}} \right\rfloor$. Then $N = k^d$ implies

$$\frac{1}{2^d}\left(\frac{1}{2\kappa}\right)^{\frac{d}{p\alpha+d}} \leq N \leq \left(\frac{1}{2\kappa}\right)^{\frac{d}{p\alpha+d}}.$$

This yields a bound of

$$\varepsilon^{(p)}_{\geq\kappa}(W) \leq 4D_1\left(2^d(2\kappa)^{\frac{d}{p\alpha+d}}\right)^{\frac{\beta'\alpha}{d(1+\beta')}},$$

which is $O\big(\kappa^{\alpha'}\big)$. Again the implicit constant depends only on α, β, p, C, and π. □

APPENDIX G: POWER-LAW GRAPHS

Recall that the normalized degree distribution of a graph G on $[n]$ is defined as the empirical distribution of the normalized degrees $d_i/\bar{d}$, where $\bar{d}$ is the average degree. We say that a sequence $(G_n)_{n\geq 0}$ has *convergent degree sequences* if the cumulative distribution functions D_{G_n} of the normalized degrees converge to some distribution function D in the Lévy-Prokhorov distance d_{LP} or, equivalently, if $D_{G_n}(\lambda) \to D(\lambda)$ for all λ at which D is continuous. (Recall that a *distribution function* is a nondecreasing, right-continuous function $D\colon \mathbb{R} \to [0,1]$ such that $\lim_{\lambda\to-\infty} D(\lambda) = 0$ and $\lim_{\lambda\to\infty} D(\lambda) = 1$.)

We say that the sequence $(G_n)_{n\geq 0}$ has a *power-law degree distribution with exponent* γ if its degree distributions converge to D satisfying

$$D(\lambda) = 1 - \Theta\big(\lambda^{-(\gamma-1)}\big) \qquad \text{as } \lambda \to \infty,$$

and we say that a graphon W has a *power-law degree distribution with exponent* γ if $D_W = 1 - \Theta\big(\lambda^{-(\gamma-1)}\big)$ as $\lambda \to \infty$.

Note that it is $\gamma - 1$ that appears in the exponent, not γ. The naming conventions in the above definitions are based on density functions, rather than distribution functions: if the degree distribution is absolutely continuous with respect to Lebesgue measure and thus has a density function $f(\lambda)$, and if $f(\lambda) = \Theta\big(\lambda^{-\gamma}\big)$ as $\lambda \to \infty$, then the distribution function D satisfies

$$1 - D(\lambda) = \int_\lambda^\infty f(\lambda)\, d\lambda = \Theta\big(\lambda^{-(\gamma-1)}\big).$$

In this appendix, we give two examples of W-random graphs with power-law degree distributions and establish bounds on the convergence rate of our estimation procedures for these graphons.

We start with an example that can be expressed as a Hölder-continuous graphon over $\mathbb{R}^d$, even though we will first define it as a graphon over $[0,1]$. It is the graphon

$$W(x,y) = \frac{1}{2}(g(x) + g(y)) \quad \text{where } g(x) = (1-\alpha)(1-x)^{-\alpha}. \tag{G.1}$$

for some $\alpha \in (0,1)$. Note that the degrees of this graphon are $W_x = \frac{1}{2} + \frac{1}{2}g(x)$, with a distribution function $D_W(\lambda)$ that is $1 - \Theta\big(\lambda^{-1/\alpha}\big)$ as $\lambda \to \infty$, showing that the graphs $G_n(\rho_n W)$ have a power-law degree distribution with exponent $\gamma = 1 + \frac{1}{\alpha}$.

As a graphon over $[0,1]$ equipped with the uniform measure, this graphon is not continuous, but it turns out that it can be expressed as an equivalent graphon over $\mathbb{R}^d$ that is Hölder-continuous. To see this, let us consider a probability distribution π on $\mathbb{R}^d$ such that the distribution of the L^2 norm $r = |x|_2$ of $x \in \mathbb{R}^d$ is absolutely continuous with respect to the Lebesgue measure on $[0,\infty)$, with a strictly positive density function $f(r)$. We will want to construct a measure-preserving map $\phi\colon \mathbb{R}^d \to [0,1)$ to obtain an equivalent graphon W^ϕ over $\mathbb{R}^d$. Requiring ϕ to be measure preserving is equivalent to requiring that $\pi(\phi^{-1}([0,a])) = \pi(\{x : \phi(x) \leq a\}) = a$. We will construct ϕ radially, via a map F such that $\phi(x) = F(|x|_2)$, and we will make sure that F is strictly increasing, in which case $\phi(x) \leq a$ is equivalent to $|x|_2 \leq F^{-1}(a)$. Thus, our condition for ϕ to be measure preserving becomes

$a = \int 1_{|x|_2 \le F^{-1}(a)}\, d\pi(x)$, or equivalently, $\int 1_{|x|_2 \le r}\, d\pi(x) = F(r)$, showing that $F(r)$ is the cumulative distribution function of $|x|_2$ (which is strictly monotone by our assumption that $f(r) > 0$ for all $r \in [0, \infty)$). Taking $F(r) = 1 - \frac{1}{r+1}$, we get

$$\begin{aligned} W^\phi(x,y) &= \frac{1-\alpha}{2}\Big(\frac{1}{1-F(|x|_2)}\Big)^\alpha + \frac{1-\alpha}{2}\Big(\frac{1}{1-F(|y|_2)}\Big)^\alpha \\ &= \frac{1-\alpha}{2}\big((1+|x|_2)^\alpha + (1+|y|_2)^\alpha\big), \end{aligned}$$

showing that W is equivalent to an α-Hölder-continuous graphon over $\mathbb{R}^d$ equipped with any measure for which the cumulative distribution function of $|x|_2$ is equal to F. As a consequence, we may use the results of Section 7 to give explicit bounds on the estimation errors for the least squares and least cut algorithms. We will not give these bounds here, since for W of the form (G.1), one can obtain slightly better bounds using the actual form of W; see Lemma G.1 below.

The second example we consider in this appendix is the graphon W over $[0,1]$ that is defined by

$$W(x,y) = g(x)g(y) \quad \text{where again } g(x) = (1-\alpha)(1-x)^{-\alpha}. \tag{G.2}$$

As before, we equip $[0,1]$ with the uniform measure. Now the degrees of W are equal to $g(x)$, which shows that again, the W-random graphs obtained from W have power-law degrees with exponent $\gamma = 1 + \frac{1}{\alpha}$.

Note that the second graphon cannot be expressed as a Hölder-continuous graphon over $\mathbb{R}^d$ in the sense of Section 7. Indeed, suppose $\tilde{W}$ were such a graphon. By Theorem A.5, there would exist a standard Borel twin-free graphon U such that $\tilde{W} = U^\phi$ for some measure-preserving map ϕ from $\mathbb{R}^d$ to the space on which U is defined. Since W is twin-free as well we may without loss of generality assume that $U = W$ (use Theorem A.4). But this means that $\tilde{W}$ would be of the form $\tilde{W}(x,y) = W(\phi(x), \phi(y)) = g(\phi(x))g(\phi(y))$ for some measure-preserving map $\phi\colon \mathbb{R}^d \to [0,1]$. Since $g(\phi(x))$ is unbounded, this cannot be a Hölder-continuous function of the argument y.

Nevertheless, we can give explicit bounds on our estimation error since for W of the form (G.1) or (G.2), we can estimate $\varepsilon^{(p)}_{\ge\kappa}(W)$ and $\mathrm{tail}^{(p)}_\rho(W)$ directly.

LEMMA G.1. *Let $\alpha \in (0,1)$, let $1 \le p < 1/\alpha$, and define $\alpha' = \frac{1}{p} - \alpha$ and $\beta' = \frac{1-p\alpha}{p\alpha}$. If W is the power-law graphon (G.1), then*

$$\varepsilon^{(p)}_{\ge\kappa}(W) = O\big(\kappa^{\alpha'}\big) \qquad \textit{and} \qquad \mathrm{tail}^{(p)}_\rho(W) = O\big(\rho^{\beta'}\big),$$

while if W is the power-law graphon (G.2), *then*

$$\varepsilon^{(p)}_{\geq\kappa}(W) = O\big(\kappa^{\alpha'}\big) \qquad \textit{and} \qquad \mathrm{tail}^{(p)}_{\rho}(W) = O\big(\rho^{\beta'}|\log\rho|^{1/p}\big).$$

Proof. We start with the proof of the tail bounds. Defining functions $g_1, g_2\colon [0,1]^2 \to [0,\infty)$ by $g_1(x,y) = g(x)$ and $g_2(x,y) = g(y)$, we write the first graphon as $\frac{1}{2}(g_1+g_2)$. Noting that $W \geq \rho^{-1}$ implies that either $g_1 \geq 1/\rho$ or $g_2 \geq 1/\rho$, we bound

$$\begin{aligned}\|W - W_\rho\|_p &\leq \|W 1_{W\geq 1/\rho}\|_p \leq \|W(1_{\rho g_1\geq 1} + 1_{\rho g_1 \geq 1})\|_p\\ &= \frac{1}{2}\|g_1 1_{\rho g_1\geq 1} + g_2 1_{\rho g_1\geq 1} + g_1 1_{\rho g_2\geq 1} + g_2 1_{\rho g_2\geq 1}\|_p\\ &\leq \|g 1_{\rho g\geq 1}\|_p + \|1_{\rho g \geq 1}\|_p.\end{aligned}$$

The two terms can easily be calculated explicitly, giving a term of order $O\big(\rho^{\frac{1-p\alpha}{p\alpha}}\big)$ for the first and a term of order $O\big(\rho^{\frac{1}{p\alpha}}\big)$ for the second. For the second graphon, we note that the condition $W(x,y) \geq 1/\rho$ is equivalent to $(1-x)(1-y) \leq \big(\rho(1-\alpha)^2\big)^{1/\alpha}$. Changing to the variables $1-x$ and $1-y$, we have to estimate the integral

$$\int_0^1\int_0^1 (xy)^{-p\alpha} 1_{xy\leq\rho^{1/\alpha}}\, dx\, dy.$$

The integral can again be calculated explicitly, giving an error term of order $O\big(\rho^{\frac{1-p\alpha}{\alpha}}|\log\rho|\big)$. Taking the p^{th} root, we obtain the claimed tail bound for the second graphon.

All that remains is to estimate the oracle errors. Let $I_1, \dots, I_k$ be a partition of $[0,1]$ into k adjacent intervals of size $\varepsilon = \frac{1}{k}$ (ordered from left to right), let g' be the function obtained by averaging g over these intervals on $I_1 \cup I_2 \dots \cup I_{k_0}$ (where k_0 will be determined later), and let $g' = 0$ on the remaining intervals. Define $g_1, g_2\colon [0,1]^2 \to [0,\infty)$ as above, define g'_1 and g'_2 analogously, and set $W' = \frac{1}{2}(g'_1 + g'_2)$ for the graphon (G.1) and $W' = g'_1 g'_2$ for the graphon (G.2). With this notation,

$$\|W - W'\|_p = \frac{1}{2}\|g_1 + g_2 - g'_1 - g'_2\|_p = \|g - g'\|_p$$

for the graphon (G.1), and

$$\|W - W'\|_p = \|g_1 g_2 - g'_1 g'_2\|_p \leq \|(g_1 - g'_1)g_2\|_p + \|g'_1(g_2 - g'_2)\|_p \leq \|g\|_p\|g - g'\|_p$$

for the graphon (G.2). So all we need to show is that $\|g - g'\|_p = O\big(\varepsilon^{\alpha'}\big)$.

For $i \le k_0$, let $\bar{x}_i \in I_i$ be defined by $\frac{1}{\varepsilon}\int_{I_i} g = g(\bar{x}_i)$. For $x \in I_i$, we bound $|g(x) - g(\bar{x}_i)| \le \max_{y \in I_i}\left|\frac{dg(y)}{dy}\right||x - \bar{x}_i|$, implying that the integral of $|g(x) - g(\bar{x}_i)|^p$ over I_i can be bounded by $\varepsilon^{p+1}\max_{y\in I_i}\left|\frac{dg(y)}{dy}\right|^p \le \varepsilon^{p+1}(1-i\varepsilon)^{-p(1+\alpha)}$. Summing over $i = 1, \dots, k_0$, we get a contribution of $O\big(\varepsilon^p(1-k_0\varepsilon)^{1-p(1+\alpha)}\big)$ to $\|g-g'\|_p^p$. The integral of g^p from $k_0\varepsilon$ to 1 will contribute $O\big((1-k_0\varepsilon)^{1-\alpha p}\big)$. As a consequence, the choice $k_0 = k-1$ (which yields $1 - k_0\varepsilon = \varepsilon$) leads to the estimate

$$\|g - g'\|_p^p = O\big(\varepsilon^{1-\alpha p}\big),$$

as desired. $\square$

APPENDIX H: CONCENTRATION BOUNDS

We start with a slight generalization of the multiplicative Chernoff bound.

LEMMA H.1. *Let $X_1, \dots, X_n$ be independent random variables with values in $\mathbb{R}$, let $X = \sum_{i=1}^n X_i$, and suppose there exists $X_0 \in [0, \infty)$ such that*

$$\sum_i \mathbb{E}[X_i^m] \le X_0 \quad \textit{for all } m \ge 2.$$

Then

$$\Pr(X - \mathbb{E}[X] \ge X_0 t) \le \exp\left(-\min\{t, t^2\}\frac{X_0}{3}\right)$$

for $t \ge 0$.

PROOF. As in the proof of the standard Chernoff bound, we estimate the expectation of $e^{\alpha X}$ for a constant $\alpha \ge 0$ to be determined later. To this end, we first bound

$$\begin{aligned}\mathbb{E}[e^{\alpha X_i}] &= 1 + \alpha\mathbb{E}[X_i] + \sum_{m\ge 2}\frac{\alpha^m \mathbb{E}[X_i^m]}{m!}\\ &\le \exp\Big(\alpha\mathbb{E}[X_i] + \sum_{m\ge 2}\frac{\alpha^m\mathbb{E}[X_i^m]}{m!}\Big),\end{aligned}$$

which together with the assumption of the lemma proves that

$$\mathbb{E}[e^{\alpha X}] \le \exp\Big(\alpha\mathbb{E}[X] + \sum_{m\ge 2}\frac{\alpha^m}{m!}\sum_i \mathbb{E}[X_i^m]\Big) \le e^{\alpha\mathbb{E}[X] + (e^\alpha - \alpha - 1)X_0}.$$

As a consequence,

$$\begin{aligned}\Pr\bigl(X \geq E(X) + tX_0\bigr) &= \Pr\Bigl(e^{\alpha X - \alpha\mathbb{E}[X] - t\alpha X_0} \geq 1\Bigr)\\ &\leq \mathbb{E}[e^{\alpha X}]e^{-\alpha\mathbb{E}[X] - t\alpha X_0}\\ &\leq e^{(e^\alpha - \alpha - 1)X_0 - t\alpha X_0}.\end{aligned}$$

Choosing $\alpha = \log(1+t)$ yields $e^\alpha - 1 - \alpha - t\alpha = t - (t+1)\log(t+1)$ and hence

$$\Pr\bigl(X \geq E(X) + tX_0\bigr) \leq e^{-X_0((t+1)\log(t+1) - t)} \leq \exp\left(-\frac{X_0}{3}\min\{t, t^2\}\right).$$

$\square$

Lemma H.1 immediately implies the following lemma. To state it, we define, for an arbitrary symmetric matrix $Q \in [0,1]^{n\times n}$ with empty diagonal, the random symmetric matrix $A = \mathrm{Bern}(Q) \in \{0,1\}^{n\times n}$ obtained by setting $A_{ij} = A_{ji} = 1$ with probability Q_{ij}, independently for all $i < j$, and $A_{ij} = 0$ whenever $i = j$. Note that with this notation, $\mathbb{E}[A_\tau] = Q_\tau$ for all $\tau\colon [n] \to [k]$. The following lemma states that A_τ is tightly concentrated around its expectation.

Lemma H.2. *Let $1 \leq k \leq n$, let Q be a symmetric $n \times n$ matrix with entries in $[0,1]$ and empty diagonal, and let $A = \mathrm{Bern}(Q)$. Let ε be the random variable $\varepsilon = \max_{\tau\colon [n]\to[k]} \|A_\tau - Q_\tau\|_1$. Then*

$$\mathbb{E}[\varepsilon] \leq 9\sqrt{\rho(Q)\left(\frac{1+\log k}{n} + \frac{k^2}{n^2}\right)}. \tag{H.1}$$

If $n\rho(Q) \geq 1$, then with probability at least $1 - e^{-n}$

$$\varepsilon \leq 8\sqrt{\rho(Q)\left(\frac{1+\log k}{n} + \frac{k^2}{n^2}\right)}. \tag{H.2}$$

Recall that $\rho(Q)$ means $\frac{1}{n^2}\sum_{i,j} Q_{ij}$.

Proof. We begin with the proof of (H.2). We distinguish two cases:

If $\frac{1+\log k}{n} + \frac{k^2}{n^2} \geq \rho(Q)$, all we need to show is that with probability at least $1 - e^{-n}$, the left side is at most $8\rho(Q)$. To prove this, we bound

$$\|A_\tau - Q_\tau\|_1 \leq \|A_\tau\|_1 + \|Q_\tau\|_1 = \|A\|_1 + \|Q\|_1.$$

Now we apply Lemma H.1 to the random variable $X = \sum_{i<j} A_{ij}$. Because $\mathbb{E}[\sum_{i<j} A_{ij}^m] = \sum_{i<j} Q_{ij} = \frac{n^2}{2}\rho(Q)$, we can take $X_0 = \frac{n^2}{2}\rho(Q)$. Taking $t = 6$, we see that with probability at least $1 - e^{-n^2\rho(Q)} \geq 1 - e^{-n}$,

$$\|A_\tau - Q_\tau\|_1 \leq 2\|Q\|_1 + 6\rho(Q) = 8\rho(Q).$$

If $\frac{1+\log k}{n} + \frac{k^2}{n^2} \leq \rho(Q)$, we will use a union bound over all $\tau\colon [n] \to [k]$. Considering first a fixed $\tau\colon [n] \to [k]$, we rewrite

$$\begin{aligned}\|A_\tau - Q_\tau\|_1 &= \frac{2}{n^2}\sum_{u<v}(Q_{uv} - A_{uv})\operatorname{sign}((Q_\tau)_{uv} - (A_\tau)_{uv}) \\ &= \max_{B\in\mathcal{B}_\tau} \frac{2}{n^2}\sum_{u<v} B_{uv}(Q_{uv} - A_{uv}),\end{aligned}$$

where $\mathcal{B}_\tau$ consists of all matrices with entries ± 1 that are constant on the partition classes of τ (note that $\mathcal{B}_\tau$ has size $2^{k_0^2}$, where $k_0 \leq k$ is the number of non-empty classes in τ). Applying Lemma H.1 again, this time to the random variables $B_{uv}A_{uv}$, noting that $\sum_{u<v} \mathbb{E}[(B_{uv}A_{uv})^\alpha] \leq \sum_{u<v} \mathbb{E}[A_{uv}] \leq \frac{n^2}{2}\rho(Q)$, and using the union bound to deal with the maximum over $B' \in \mathcal{B}_\tau$, we find that

$$\Pr\left(\|A_\tau - Q_\tau\|_1 \geq t\rho(Q)\right) \leq 2^{k^2} \exp\left(-\frac{\min\{t, t^2\}}{6} n^2\rho(Q)\right).$$

Setting

$$t = 6\sqrt{\frac{1+\log k}{n\rho(Q)} + \frac{k^2}{n^2\rho(Q)}},$$

our case assumption implies that $t \leq 6$, which in turn implies that

$$\min\{t, t^2\} \geq \frac{t^2}{6} = 6\left(\frac{1+\log k}{n\rho(Q)} + \frac{k^2}{n^2\rho(Q)}\right).$$

As a consequence, for each partition $\tau\colon [n] \to [k]$,

$$\begin{aligned}\Pr(\|A_\tau - Q_\tau\|_1 \geq t\rho(Q)) &\leq \exp\left(k^2\log 2 - n(1+\log k) - k^2\right) \\ &\leq e^{-n(1+\log k)}.\end{aligned}$$

Taking the union bound over all partitions $\tau\colon [n] \to [k]$, we obtain the desired bound.

All that remains is to prove (H.1). If $n\rho(Q) \leq 1$, we bound

$$\mathbb{E}[\varepsilon] \leq \mathbb{E}[\|A - Q\|_1] \leq \|Q\|_1 + \mathbb{E}[\|A\|_1] = 2\rho(Q) \leq 2\sqrt{\rho(Q)/n}.$$

If $n\rho(Q) \geq 1$, we use (H.2) and the fact that $\varepsilon \leq \|Q\|_1 + \|A\|_1 \leq 2$ to bound

$$\mathbb{E}[\varepsilon] \leq 8\sqrt{\rho(Q)\left(\frac{1+\log k}{n} + \frac{k^2}{n^2}\right)} + 2e^{-n}.$$

Because $2e^{-n} \leq 1/n \leq \sqrt{n\rho(G)}/n = \sqrt{\rho(G)/n}$, this completes the proof. □

Our next lemma states that a similar bound holds for the cut norm of $A - Q$.

Lemma H.3. *Let $n \geq 2$, let Q be a symmetric $n \times n$ matrix with entries in $[0, 1]$ and empty diagonal, and let $A = \mathrm{Bern}(Q)$. Then*

$$\mathbb{E}[\|A - Q\|_\square] \leq 16\sqrt{\frac{\rho(Q)}{n}}.$$

If $n\rho(Q) \geq 1$, then with probability at least $1 - e^{-n}$,

$$\|A - Q\|_\square \leq 15\sqrt{\frac{\rho(Q)}{n}}. \tag{H.3}$$

Proof. A bound of the form (H.3) can easily be inferred from Lemma 7.2 in [2]. For the convenience of the reader, we given an independent, slightly simpler proof here.

Let $\mathcal{F}_n$ be the set of functions $f\colon [n] \to \{-1, +1\}$. It is not hard to check that

$$\begin{aligned}\|A - Q\|_\square &\leq \max_{f,g\in\mathcal{F}_n} \frac{1}{n^2}\sum_{i,j} f(i)g(j)(A_{ij} - Q_{ij})\\ &\leq \max_{f,g\in\mathcal{F}_n} \frac{2}{n^2}\sum_{i<j} f(i)g(j)(A_{ij} - Q_{ij}).\end{aligned}$$

Proceeding as in the proof of Lemma H.2, a union bound and Lemma H.1 now imply that

$$\Pr\left(\|A - Q\|_\square \geq t\rho(Q)\right) \leq 4^n \exp\left(-\frac{\min\{t, t^2\}}{6}n^2\rho(Q)\right).$$

Choosing $t = 6\log(4e)/\sqrt{n\rho(Q)}$ and observing that $6\log(4e) \leq 15$ then gives the high probability bound. The bound in expectation follows from this bound and the observation that $\|A - Q\|_\square \leq 2\rho(Q)$. Indeed, if $n\rho(Q) \geq 1$, then

$$\begin{aligned} 15\sqrt{\rho(Q)/n} + 2e^{-n}\rho(Q) &\leq 15\sqrt{\rho(Q)/n} + 2\rho(Q)/(en) \\ &\leq 16\sqrt{\rho(Q)/n} \end{aligned}$$

(for the final step recall that $\rho(Q) \leq 1$), and if $n\rho(Q) \leq 1$, then $2\rho(Q) \leq 2\sqrt{\rho(Q)/n}$. □

REFERENCES

[1] Bollobás, B. and Riordan, O. (2009). Metrics for sparse graphs. In *Surveys in combinatorics 2009* (S. Huczynska, J. D. Mitchell and C. M. Roney-Dougal, eds.). *London Math. Soc. Lecture Note Ser.* **365** 211–287. Cambridge University Press.

[2] Borgs, C., Chayes, J. T., Cohn, H. and Zhao, Y. (2019). An L^p theory of sparse graph convergence I: Limits, sparse random graph models, and power law distributions. *Trans. Amer. Math. Soc.* **372** 3019–3062.

[3] Borgs, C., Chayes, J. and Lovász, L. (2010). Moments of two-variable functions and the uniqueness of graph limits. *Geom. Funct. Anal.* **19** 1597–1619.

[4] Borgs, C., Chayes, J. T., Lovász, L., Sós, V. and Vesztergombi, K. (2008). Convergent graph sequences I: subgraph frequencies, metric properties, and testing. *Advances in Math.* **219** 1801–1851.

[5] Borgs, C., Chayes, J. T. and Smith, A. (2015). Private graphon estimation for sparse graphs. In *Advances in Neural Information Processing Systems 28* 1369–1377. Extended version with proofs available at arXiv:1506.06162.

[6] Hoeffding, W. (1961). The strong law of large numbers for U-statistics. *North Carolina State University, Institute of Statistics Mimeograph Series* **302**. http://repository.lib.ncsu.edu/dr/handle/1840.4/2128.

[7] Janson, S. (2013). *Graphons, cut norm and distance, couplings and rearrangements. NYJM Monographs* **4**. State University of New York, University at Albany, Albany, NY. 76 pp.

C. Borgs
J. T. Chayes
S. Ganguly
University of California, Berkeley
Berkeley, CA 94720, USA
E-mail: borgs@berkeley.edu
jchayes@berkeley.edu
sganguly@berkeley.edu

H. Cohn
Microsoft Research
One Memorial Drive
Cambridge, MA 02142, USA
E-mail: cohn@microsoft.com